

MINISTRY OF HIGHER EDUCATION, SCIENCE AND
INNOVATIONS OF THE REPUBLIC OF UZBEKISTAN

NATIONAL UNIVERSITY OF UZBEKISTAN NAMED AFTER
MIRZO ULUGBEK



RAQAMLI TEXNOLOGIYALAR VA
SUN'YI INTELLEKTNI RIVOJLANTIRISH
ILMIY-TADQIQOT INSTITUTI



INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE

PROCEEDINGS
of materials on topic:

«CONTEMPORARY TECHNOLOGIES OF
COMPUTATIONAL LINGUISTICS - CTCL.2025»



Crossref
we use DOIs



Google
scholar

OPEN



ACCESS



TASHKENT - 2025

**“KOMPYUTER LINGVISTIKASINING ZAMONAVIY TEXNOLOGIYALARI – CTCL.2025”
NOMLI XALQARO ILMIY-AMALIY KONFERENSIYASINING ILMIY
MAQOLALARI VA TEZISLAR TO‘PLAMI. T.: O‘zMU, 2025.
2025-yil 25-aprel**

TASHKILIY QO‘MITA TARKIBI:

<i>N.Z.Abduraxmonova.</i>	O‘zMU Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası mudiri, f.f.d., professor.
<i>G.B.Madiyeva.</i>	Al-Farobiy nomidagi Qozoq Milliy universiteti professori, f.f.d.
<i>Z. T.Taxirov.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası professori, f.f.n.
<i>A.R.Iskandarova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası dotsenti, PhD.
<i>D.A.Fattaxova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi, PhD.
<i>U.R.Xalmuxamedova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi.
<i>N.Sh.Inogamova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi.
<i>F.R.Topildiyeva.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi.
<i>D.K.Xudoyberganova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi, PhD.
<i>Z.M.Kasimova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi, PhD.
<i>J.N.O‘rozov.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası katta o‘qituvchisi, PhD.
<i>G.G‘.Xoliyorova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
<i>F.M.Aliyev.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
<i>M.S.Jo‘raqulova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
<i>Z.G.Qodirova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
<i>M.F.Abdunazarova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
<i>D.Xolbekova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
<i>D.A.Mirvaliyeva.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
<i>N.K.Tursunova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
<i>U.M.Askarova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
<i>S.A.Abduraxmonova.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası o‘qituvchisi.
<i>R.T.O‘rinboyeva.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası magistranti.
<i>L.A.Boboxo‘jayeva.</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası magistranti.
<i>M.Sh.Xudoyberganova</i>	Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası magistranti.

Mazkur xalqaro ilmiy-amaliy konferensiya to'plami kompyuter lingvistikasining ayrim yo'nalishlari, jumladan, korpus lingvistikasi, til ta'limida kompyuter texnologiyalari, NLPda til va nutq tahlili (lingvistik analizator: morfologik, sintaktik va semantik tahlil, nutq analizi va sintezi), mashina tarjimasining lingvistik va dasturiy ta'minoti, kompyuter leksikografiyasi: ontologiya, tezaurus, shuningdek, zamonaviy tilshunoslik masalalari kabi mavzular tahlilga tortilgan maqola va tezislardan tashkil topgan.

Ushbu ilmiy maqola va tezislar to'plami kompyuter lingvistikasi, kompyuter texnologiyalari, amaliy filologiya kabi yo'nalishlarda ilmiy izlanishlar olib borayotgan professor-o'qituvchilar, doktorantlar, mustaqil izlanuvchilar, magistr va talabalar uchun mo'ljallangan.

To'plamga kiritilgan maqolalar va tezislar mazmuni hamda sifatiga mualliflar mas'uldir.

MAS'UL MUHARRIR:

prof., f.f.d. N.Z.Abduraxmonova

TAHRIR HAY'ATI:

f.f.n., prof. Z. T.Taxirov

f.f.n., dots. A.R.Iskandarova

ped.f.n., dots. D.A.Fattaxova

PhD D.K.Xudoyberganova

PhD Z.M.Kasimova

**СБОРНИК НАУЧНЫХ СТАТЕЙ И ТЕЗИСОВ
МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ
“СОВРЕМЕННЫЕ ТЕХНОЛОГИИ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ – CTCL.2025”.**
Т.: НУУз, 2025.
25 апреля 2025 г.

СОСТАВ ОРГАНИЗАЦИОННОГО КОМИТЕТА:

Н.З. Абдурахмонова.	Заведующий кафедрой компьютерной и прикладной лингвистики НУУз, д.ф.н., профессор.
Г.Б.Мадиева.	Профессор Казахского Национального университета имени аль-Фараби, д.ф.н.
З. Т. Тахиров.	Профессор кафедры компьютерной и прикладной лингвистики, к.ф.н.
А.Р.Искандарова.	Доцент кафедры компьютерной и прикладной лингвистики, к.ф.н.
Д.А. Фаттахова.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики, к.пед.н.
У.Р. Халмухамедова.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики.
Н. Ш. Иногамова.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики.
Ф.Р. Топилдиева.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики.
Д.К.Худойберганаева.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики, к.ф.н.
З.М. Касимова.	Старшая преподавательница кафедры компьютерной и прикладной лингвистики, к.ф.н.
Ж.Н.Урозов.	Преподавательница кафедры компьютерной и прикладной лингвистики, к.ф.н.
Г.Г. Холёрова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
Ф.М. Алиев.	Преподаватель кафедры компьютерной и прикладной лингвистики.
М.С. Джуракулова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
З.Г. Кадирова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
М.Ф. Абдуназарова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
Д. Холбекова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
Д.А. Мирвалиева.	Преподавательница кафедры компьютерной и прикладной лингвистики.
К.Р. Ибодуллаева.	Преподавательница кафедры компьютерной и прикладной лингвистики.
Н.К. Турсунова.	Преподаватель кафедры компьютерной и прикладной лингвистики.
У.М. Аскарова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
С.А. Абдурахмонова.	Преподавательница кафедры компьютерной и прикладной лингвистики.
Р.Т.Уринбоева.	Магистрант кафедрой компьютерной и прикладной лингвистики.
Л.А.Бобоходжаева.	Магистрант кафедрой компьютерной и прикладной лингвистики.
М.Ш.Худойберганаева	Магистрант кафедрой компьютерной и прикладной лингвистики.

Материалы международной научно-практической конференции, включающие статьи и тезисы, посвящены некоторым направлениям компьютерной лингвистики, в том числе корпусной лингвистике, компьютерным технологиям в языковом образовании, анализу языка и речи в NLP (Лингвистический анализатор: морфологический, синтаксический и семантический анализ, анализ речи и синтез), Лингвистические базы данных и программное обеспечение машинного перевода, Компьютерная лексикография: онтология, тезаурус, а также проблемы современной лингвистики.

Сборник научных статей и диссертаций предназначен для профессоров и преподавателей, докторантов, независимых исследователей, магистров и студентов, ведущих научные исследования в таких областях, как компьютерная лингвистика, компьютерные технологии и прикладная лингвистика.

Авторы несут ответственность за содержание и качество статей и тезисов, включенных в книгу.

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

д.ф.н., проф. Н.Абдурахмонова

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

PhD, проф. З.Тахиров

PhD, доц. А.Искандарова

PhD З.Касимова

PhD Д.Фаттахова

PhD Д.Худойбергана

**PROCEEDINGS OF ARTICLES AND THESIS
 OF THE INTERNATIONAL SCIENTIFIC -PRACTICAL CONFERENCE TITLED
 «CONTEMPORARY TECHNOLOGIES OF COMPUTATIONAL LINGUISTICS-CTCL.2025»
 T.: NUUz, 2025.
 April 25, 2025.**

ORGANIZING COMMITTEE:

<i>N. Z. Abdurakhmonova.</i>	Head of Department of Computational and Applied Linguistics of UzMU, DSc, Professor.
<i>G.B.Madiev.</i>	Professor of the Al-Farabi Kazakh National University, Doctor of Philological Sciences
<i>Z. T. Takhirov.</i>	Professor of the Department of Computational and Applied Linguistics, PhD.
<i>A.R. Iskandarov.</i>	Associate Professor of the Department of Computational and Applied Linguistics, PhD.
<i>D.A. Fattakhova.</i>	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
<i>U.R. Khalmukhamedova.</i>	Senior teacher of the Department of Computational and Applied Linguistics.
<i>N. Sh. Inogamova.</i>	Senior teacher of the Department of Computational and Applied Linguistics.
<i>F.R. Topildieva.</i>	Senior teacher of the Department of Computational and Applied Linguistics.
<i>D.K. Khudoyberganova.</i>	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
<i>Z.M. Kasimova.</i>	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
<i>J.N.Urazov.</i>	Senior teacher of the Department of Computational and Applied Linguistics, PhD.
<i>G.G. Kholiyorova</i>	Senior teacher of the Department of Computational and Applied Linguistics.
<i>F.M. Aliyev.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>M.S. Jurakulova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>Z.G. Kadirova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>M.F. Abdunazarova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>D. Kholbekova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>D.A. Mirvaliyeva.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>N.K.Tursunova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>U.M. Askarova.</i>	Teacher of the Department of Computational and Applied Linguistics.
<i>S.A. Abdurakhmonova.</i>	Teacher of the Department of Computational and Applied Linguistics
<i>R.T.Urinboeva.</i>	Master student Department of Computational and Applied Linguistics.
<i>L.A.Bobokhodzhaeva.</i>	Master student Department of Computational and Applied Linguistics.
<i>M.Sh.Khudoyberganova</i>	Master student Department of Computational and Applied Linguistics

This proceedings of international scientific-practical conference including articles and theses deals with some directions of Computational Linguistics, including Corpus Linguistics, Computer Technologies in language education, Language and speech analysis in NLP (Linguistic analyzer: morphological, syntactic and semantic analysis, speech analysis and synthesis), Linguistic database and software of Machine Translation, Computer Lexicography: ontology, thesaurus, as well as Modern Linguistic Issues.

The proceedings of scientific articles and theses is intended for professors and teachers, doctoral students, independent researchers, masters and students who are conducting scientific research in such areas as Computational Linguistics, Computer Technologies, and Applied Linguistics.

The authors are responsible for the content and quality of the articles and theses included in the book.

RESPONSIBLE EDITOR:

DSc, prof. N.Abdurakhmonova

EDITORIAL BOARD:

PhD, prof. Z.Taxirov

PhD, dots. A.Iskandarova

PhD Z.Kasimova

PhD D.Fattaxova

PhD D.Xudoyberganova

MUNDARIJA

TABLE OF CONTENTS

N. ABDURAXMONOVA, U. TULIYEV, G. SHAMIYEVA. PARATRANSLATOR – PARALLEL KORPUSGA ASOSLANGAN ELEKTRON KONTEKSTOLOGIK TARJIMA LUG‘AT PLATFORMASINI YARATISHNING AYRIM MASALALARI XUSUSIDA	15
D. ZHUPARBEEK, S.S. NURKENOVA. THE USE OF WEBQUESTS IN FOREIGN LANGUAGE TEACHING TO FOSTER STUDENTS’ COMMUNICATIVE COMPETENCE	21
M. AMIRSHEIBANI, O.A. YO‘LDOSHEV. A CORPUS-BASED DESCRIPTION OF GRAMMATICAL ERRORS IN UZBEK EFL LEARNER WRITING	26
Н.А. ИСМАЙЛОВА. КОМПЬЮТЕРЛІК ЛИНГВИСТИКА ЖӘНЕ ЖАСАНДЫ ИНТЕЛЛЕКТ: ТІЛДІК КЕДЕРГІЛЕРДІ ЖЕҢУДІҢ ЗАМАНАУИ ТӘСІЛДЕРІ	33
Д. ТОҚМЫРЗАЕВ, А. ПИРМАНОВА. СОЗДАНИЕ ПОЭТИЧЕСКОГО ЧАСТОТНОГО СЛОВАРЯ СЛОВ В СТИХАХ АБАЯ В СООТВЕТСТВИИ С КОРПУСОМ, АВТОМАТИЗИРОВАННЫЙ ЧАСТОТНЫЙ СЛОВАРЬ (1-Й ЭТАП)	35
Г. МАДИЕВА, К. ПИРМАНОВА. СОЗДАНИЕ И АНАЛИЗ ЛИТЕРАТУРНО-ЯЗЫКОВОГО КОРПУСА СТИХОТВОРЕНИЙ АБАЯ.....	40
А. ОҢАЛБЕК, А. УӘЛИХАН. ПОТЕНЦИАЛ ИСПОЛЬЗОВАНИЯ ПОЭТИЧЕСКОГО КОРПУСА АБАЯ В ИССЛЕДОВАТЕЛЬСКИХ, ОБРАЗОВАТЕЛЬНЫХ И МЕТОДИЧЕСКИХ ЦЕЛЯХ	44
M.I. SUYUNOVA. SEMANTIC AMBIGUITY PROCESSING IN HUMAN COGNITION: A COMPARATIVE ANALYSIS OF NLP MODELS AND PSYCHOLINGUISTIC APPROACHES	55
M.YA. URAZALIYEVA. MACHINE LEARNING MODELLARI ASOSIDA O‘ZBEK TILIDAGI AUDIOMATNLARNI TAHLIL QILISH: KUN.UZ MISOLIDA.....	58
S. TURSUNBOYEVA. MILLIY KONTENT YARATISHDA ADABIY TIL ME‘YORLARIGA RIOYA QILISH.....	62
N.Z. ABDURAXMONOVA, K.CH. RAJABOVA. MULTIMEDIA KORPUSLARINING TIL O‘QITISHDAGI O‘RNI VA AHAMIYATI.....	66
R.S. SHAMSIYEVA. CHET ELLIK TALABALARNI O‘QITISHDA KORPUSGA ASOSLANGAN YONDASHUV	71
A.Q. ABDULLAYEV. SENTIMENT TAHLILDA TASNIFLASH SXEMASI	75
N.Z. ABDURAXMONOVA, D.G. ALLABERDIYEVA. QORAQALPOQ TILI ELEKTRON KORPUSINI YARATISH: NAZARIY, AMALIY VA STRATEGIK ASOSLAR	79
U.T. TURAYEVA. O‘ZBEK VA TURK TILLARIDA JOY NOMLARINING IFODALANISHI.....	83
G.SH. SAIDOVA, D.A. ABDUSATTAROVA. O‘ZBEK TILIDAGI METAFORALARNING INGLIZ TILIGA TARJIMASIDA TARJIMA MUAMMOLARI.....	89
M.S. SHARIPOV, L.U. KURBANOVA, M. ATANAZAROVA. XORAZM SHEVALARI SINONIMIK LUG‘ATINING MA‘LUMOTLAR BAZASINI ISHLAB CHIQUISH.....	93
S.A. TURSUNOVA. JAHON TILSHUNOSLIGIDA PRAGMALINGVISTIKA VA PRAGMATIK TAHLIL	96

K. IBODULLAYEVA. DISKURS TUSHUNCHASI VA UNING INTERTEKSTUALLIK HODISASI BILAN BOG‘LANISHI	100
X.U. MAXSUDOVA. MASHINASOZLIK TERMINOLOGIYASINING SHAKLLANISH BOSQICHLARI	104
M. SHAYDULLAYEVA, M. FAZLIDDINOVA, X. ISMANQULOVA, A. MAMARAYIMOVA. O‘ZBEK TILINING MORFOLOGIK ANALIZATORINI TADQIQ ETISH VA YARATISH	107
G.A. SHAMSIYEVA. MASHINA TARJIMASI UCHUN TARJIMA XOTIRASINI YARATISH TAMOYILLARI	110
У.С. ЙИГИТАЛИЕВ, Ш.У. УМИДЖОНОВ. ВЫРАЖЕНИЕ ЯЗЫКОВОГО АТОМА «РОД» НА ГРАММАТИЧЕСКОМ УРОВНЕ УЗБЕКСКОГО ЯЗЫКА.....	114
M.F. ABDUNAZAROVA. PRAGMALINGVISTIKA VA UNDA DISKURS TUSHUNCHASI	117
A. ISKANDAROVA. “QIYOMAT” ROMANIDAGI FRAZEMALARNI PARALLEL KORPUSDA LINGVISTIK TEGGLASH MASALASI	121
Z.M. KASIMOVA. SEMANTIC-STRUCTURAL ANALYSIS OF EXPRESSIONS OF THE BEGINNING OF THE XX CENTURY	125
D.D. FATTAKHOVA, D.A.MIRVALIYEVA. THE FORMATION AND DEVELOPMENT OF LINGVODIDACTICS	129
И. ДЖУМАНИЯЗОВА. ОСОБЕННОСТИ ПОСТРОЕНИЯ И ФУНКЦИОНИРОВАНИЯ АВТОРСКИХ КОПИУСОВ	137
G.G‘. XOLIYOROVA. O‘ZBEK TILI TEZAURUSI SEMANTIK MAYDONLARI VA STRUKTURASI.....	141
S.K. SA‘DULLAYEV. O‘ZBEK TILIDAGI IJTIMOIIY TARMOQ MATNLARI SENTIMENT TAHLILI	145
M.Q. JALG‘ASBEKOVA. O‘ZBEK ADABIY TILINING QOZOQ TILIGA TA‘SIRI.....	149
A.A. MOHICHEHRA, N.Z. ABDURAXMONOVA. O‘ZBEK TILINING SIFAT LEKSEMALARIDA KONTEKSTOLOGIK SINONIMLAR BAZASINI YARATISH MASALALARI.....	152
S. OTAMURODOVA. POLYSEMOUS NATURE OF HIGH-FREQUENCY PHRASAL VERBS	160
G.D. YO‘LDOSHBOYEVA. O‘ZBEK TILIDAGI ZAMONAVIIY TARJIMA TIZIMLARINI YARATISHNING DOLZARB MASALALARI	164
Г. ХИДИРОВА. ПРИМЕНЕНИЕ КОРПУСНОГО ПОДХОДА В ИССЛЕДОВАНИЯХ ЯЗЫКА	169
M.J. JO‘RABOYEVA, O. HAZRATQULOV. ZAMONAVIIY TILSHUNOSLIK VA TARJIMASHUNOSLIKNING DOLZARB MUAMMOLARI.....	173
M.B. MUXAMADKULOVA, A.R. ISKANDAROVA. O‘ZBEK VA TATAR MILLIIY KORPUSLARI TOKSONOMIYASI.....	176
R.T. O‘RINBOYEVA, U.T. TO‘RAYEVA. NUTQ SINTEZATORI UCHUN MO‘LJALLANGAN NLP PAKETLARI	181
N.R. ALOYEV. KORPUS MATNLARINI TEMATIK MODELLASHTIRISHDAGI ZAMONAVIIY YONDASHUVLAR	185

N.R. ALOYEV. ROBERTA TIL MODELI VA NMF TEMATIK MODELLASHTIRISH USULI	192
M. BOTIROVA, Z.G. QODIROVA. KORPUS UCHUN LUG‘AT TO‘PLASH VA ULARNI TASNIFLASH.....	203
N. ASROROVA. COMBINING CORPUS LINGUISTICS WITH CRITICAL DISCOURSE ANALYSIS	209
J.N. O‘ROZOV. AMALIY TILSHUNOSLIKDA LINGVISTIK EKSPERTIZA VA UNING AHAMIYATI	213
D.M. JAMOLITDINOVA, M.SH. AKBAROVA. BADIY MATN LINGVOPOETIKASIDA G‘AYRIODATIY BIRIKMALAR.....	217
K.K. MUSULMONOVA. LINGVISTIK EKSPERTIZANING KRIMINAL VAZIYATLARGA TATBIQI MASALASI.....	221
I.A. XOLMURODOVA. TEZAUROSLARDA TUR-JINS MUNOSABATI.....	224
D. XUDAYBERGANOVA. DOSTONLAR LEKSIKASIDA UCHRAYDIGAN O‘XSHATISHLI BIRLIKLARNING TIPLARI.....	229
SH. ABDULLAYEVA. O‘ZBEK VA NEMIS TILLARIDA FRAZEOLOGIK BIRLIKLARNING LINGVOKULTUROLOGIK XUSUSIYATLARI.....	233
N. SHAMIYEVA, M. JUMAYEVA, O.U. KHAZRATKULOV. SYNONYMOUS AND ANTONYMOUS SERIES OF ADJECTIVES IN THE ENGLISH THESAURUS	236
R.T. O‘RINBOYEVA, I.J. ASQAROV. SAMARQAND SHEVASI DIALEKTINI TRANSKRIPSIYA QILISH TAMOYILLARI.....	239
M. ASQAROVA. KO‘P KOMPLEMENTLI TERMINLARNI PARALLEL KORPUSDA MOSLASHTIRISH MUAMMOLARI INGLIZ VA RUS TILLARI MISOLIDA	244
B.B. ELOV, M.T. SAMATBOYEVA. NER: TEGGLASH MUOMMOLARI	247
G.B. ERKINJONOVA. SHAXS-SON QO‘SHIMCHALARINI AVTOMATIK POS-TEGLASH.....	251
SH.M. HAMROYEVA, M.X. ALIMBEKOV. “ABDURAUF FITRAT KORPUSI” MATNLARINI LEKSIK QATLAMGA KO‘RA TEGGLASH.....	255
S. NURMONOVA, N. TO‘XTABOYEVA. AVTOMATIK TARJIMADAGI MUAMMOLAR	258
R. YULDASHEVA. O‘ZBEK TILIDA KOLLOKATSIYALARGA DOIR YONDASHUVLAR	261
R. YANGIBAYEVA. DIAKRON KORPUS YARATISHNING NAZARIY VA AMALIY AHAMIYATI	266
Z. XUSAINOVA, S. YANGIBAYEVA. MUSTAQILLIK DAVRI NASHRLARIGA ASOSLANGAN DIAKRON KORPUS YARATISHNING LINGVISTIK TA‘MINOTI.....	270
M.S. SHARIPOV, O.O. SOBIROV. SAVOL-JAVOB TIZIMLARIDA LEMMALASH ALGORITMINING O‘RNI.....	274
K.T. ODILOVA. COGNITIVE FEATURES OF ADVERTISING TEXTS WITH VERBAL AND NONVERBAL COMPONENTS.....	278
L. BABAJANOVA. SEMANTIK VALENTLIK TUSHUNCHASINING YUZAGA KELISH USULLARI	294

M.S. SHARIPOV, X.S. ADINAYEV. QO‘SHMA GAPLARNI PUNKTUATSION TAHLIL QILISH ALGORITMINI ISHLAB CHIQISH.....	300
A.CH. JUMAYEVA. TIL TA’LIMIDA KOMPYUTER TEXNOLOGIYALARI.....	305
A.A. ALLASHEV. THE ROLE OF CONFLICT IN THE STRUCTURE OF A LITERARY TEXT.....	309
M. SAPAROVA. SEMANTIC CATEGORIZATION OF NOUNS IN THE UZBEK LANGUAGE THESAURUS DICTIONARY	315
N.SH. INOGAMOVA. BADIY MATN TAHLILIDA OKKAZIONALIZMLARNING LINGVOPOETIK IMKONIYATLARI	320
S.A. ABDURAXMONOVA, N.S. YASHIBOYEVA. TARJIMA MATNLARI VA ULARDA MILLIY QADRIYATLARNING TAHLILI.....	325
O‘.E. BOZORQULOVA. LINGVISTIK TADQIQOTLARDA YANGI YONDASHUV: OSKAR UAYLDNING “DORIAN GREYNING PORTRETI” HIKOYASI ASOSIDA YARATILGAN PARALLEL KORPUS VA UNING TAVSIFI.....	329
M.M. KURBANOVA, D.SH. BULTAKOVA. O‘ZBEK VA INGLIZ NUTQIY ETIKETIDA GENDERGA ASOSLANGAN KOMMUNIKATIV FARQLAR.....	334
Z.I. FAYZULLAYEVA, O.A. ABDURASHIDOVA, O.M. ZOIDJONOVA. SEMANTIK QIDIRUV ALGORITMI.....	338
A.Z. OTEMISOV. G.K. XUDAYBERGENOVA. O‘ZBEK TILI ASOSIDA QORAQALPOQ TILINING NLP TIZIMINI YARATISH	342
L.K. USMONOVA. MANIFESTATION OF THE LINGUISTIC PERSONALITY OF A POLITICIAN IN THE UZBEK LANGUAGE	345
Z.I. FAYZULLAYEVA, N.N. KARIMOV, B.M. ABDUMUTALOV, O.O. RAHMATULLAYEV. SO‘Z SHAKLLARINING ASOSINI ANIQLASH ALGORITMI	348
E.R. NAVRUZOV, M.S. JO‘RAQULOVA. MASHINALI O‘RGATISH USULLARI YORDAMIDA MATNNI SENTIMENT TAHLIL QILISH	352
M. AHMEDOVA. N-GRAM INDEKSLARIDAN FOYDALANADIGAN IMLO TUZATISH ALGORITMI.....	357
Z. QODIROVA. TEZAURUS VA ONTOLOGIYADA KONSEPT FARQLANISHI TIBBIY TERMINLAR MISOLIDA	363
E.SH. NAZIROVA, SH.B. ABIDOVA. Y.X. XOLMUMINOV. O‘ZBEK TILI GRAMATIKA SIDAGI SO‘ZLARNI TURKUMLARIGA AJRATISHDA NLP MODELLARIDAN FOYDALANILANISH	368
M.A. UMIRZAQOVA, YU.S. XOLJURAYEVA. O‘GZAKI NUTQ KOMPYUTER KORPUSLARINING AHAMIYATI VA UNI TASHKIL QILISH JARAYONI.....	376
F.R. TOPILDIYEVA. FARIDA AFRO‘Z SHE‘RIY ASARLARIDA FRAZELOGIZM VA DIALEKTIZMLAR	380
YU. XOLJURAYEVA, M.A. UMIRZAQOVA. NER TEXNOLOGIYASINING MATN TAHLILIDAGI AHAMIYATI	384
U.M. ASKAROVA. SAVODXONLIKNI OSHIRISHDA MATNLARNI AVTOMATIK TAHRIR QILISH TIZIMINING O‘RNI.....	387
D. ANVAROVA. JAHON TILSHUNOSLIGIDA OMONIMLARNING KORPUSGA ASOSLANGAN LUG‘AT BAZASINI YARATISH MASALASI.....	390

Д.А. САБИРОВА. ПРОБЛЕМА РАЗМЕТОК ПРИ ГЕНЕРАЦИИ КОРПУСА АБАЯ	394
MRS. S. JOTHI. LINGUISTIC DATABASE AND SOFTWARE OF MACHINE TRANSLATION	404
M.G'. HASANOVA. TIL TA'LIMIDA KOMPYUTER TEXNOLOGIYALARINING AHAMIYATI: CHATGPT, DUOLINGO VA IBRAT ACADEMY MISOLIDA	407
M. NORMAMATOVA, N. ABDURAXMONOVA. O'ZBEK TILINI IKKINCHI TIL SIFATIDA O'RGANISH JARAYONIDA TIL KORPUSINING ROLI VA SAMARADORLIGI	409
Г.К. МЫРЗАХАНОВА. РАЗРАБОТКА КАЗАХСКОГО ПОЭТИЧЕСКОГО КОРПУСА: МОРФОЛОГИЧЕСКАЯ И ПОЭТИЧЕСКАЯ РАЗМЕТКА СТИХОВ АБАЯ.....	413
Г. МАДИЕВА, Ж. САТКЕНОВА, Л. АЛИМТАЕВА. ЯЗЫКОВОЙ РЕСУРС ПОЭТИЧЕСКИХ КОРПУСОВ – ЦИФРОВОЙ РЕСУРС ДЛЯ АНАЛИЗА ПОЭЗИИ.....	414
M.R. SIROJIDDINOVA. BADIY ASARLAR PARALLEL KОРPUSI UCHUN SIFAT SO'Z TURKUMIGA OID FRAZEOLGIK BIRLIKЛАR BAZASINI YARATISH HAQIDA.....	419
N.Z. ABDURAXMONOVA, N.M. ABDULLAYEVA. KОРPUS UCHUN DIALEKTAL LINGVISTIK BAZA YARATISH: MUAMMO VA YECHIMLAR	424
F.B. SOLIYEVA. FONETIK VA LEKSIK JIHATIDAN BUXORO SHEVASINING TIPIK XUSUSIYATLARI	429
X. RAHIMOV. O'ZBEK TILIDAGI ADABIY MATNLARDA SENTIMENT ANALIZ IMKONIYATLARI.....	432
J.U. TIRKASHEV. O'ZBEK TILIDAGI OTLARNING CONCEPTON ASOSIDA KONSEPTUAL XARITASINI YARATISH MASALALARI.....	435
N. HAKIMOVA. O'ZBEK TILI TA'LIMIY KОРPUSI YARATILISHI, UNING TUZILISHI, IMKONIYATLARI VA KAMCHILIKLARI.....	441
N. ABDURAKHMONOVA, KH.I. ZOKIROVA. STRUCTURAL MODELS OF TRANSLATION OF UZBEK AND ENGLISH MATHEMATICAL TERMS	444
M.S. SHARIPOV, S.B. SATTAROVA. O'ZBEK TILI MATNLARINI STEMLASH ALGORITMLARI.....	450
D. OMAROVA. O'ZBEK TILIDAGI IBORALARNING SINONIMIK DARAJALANISHI	456
S.KH. SHARIPOVA. РАЗВИТИЕ ТВОРЧЕСКИХ СПОСОБНОСТЕЙ СТУДЕНТОВ НА ЗАНЯТИЯХ ПО РУССКОМУ ЯЗЫКУ И ЛИТЕРАТУРЕ В ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЯХ.....	460
M.YO'. YO'LDOSHALIYEVA. TEZAVURUS YARATISHDA MUNOSABAT VA ALOQA TUSHUNCHASI.....	466
A.Z. OTEMISOV, J.B. SHARBAEV. QARAQALPAQ HАM TУRK TILLERINI PARALLEL KОРPUSIN JARATIW MАSELESИ.....	469
A.Z. O'TEMISOV, B. MAMBETULLAEVA. QORAQOLPOQ TILIDAGI SODDA GAP TURLARINI SINTAKTIK TAHLIL BOSQICHI UCHUN TAYYORLASH.....	472
E.A. YADGAROVA. RASMIY (INSTITUTIONAL) DISKURS TUSHUNCHASI VA UNING LINGVISTIK XUSUSIYATLARI.....	476
M.I. RASHIDOVA. SINTAKTIK ANNOTATSIYALANGAN KОРPUSNING AHAMIYATI	480

N.K. TURSUNOVA. DUNYO TILSHUNOSLIGIDA SOTSIOLEKTLARNING TASNIFLANISHI VA O'RGANILISHI	483
F. TOPILDIYEVA, Z.D. RASULBERDIYEVA. BADIY MATNLARDA ARGO VA JARGONLARNING FUNKSIONAL XUSUSIYATLARI.....	487
F. MUKHAMMADALIEVA. TERMINOLOGIK TEZAVURUS YARATISHDAGI NAZARIY VA AMALIY YONDASHUVLAR BIOLOGIYA SOHASI MISOLIDA.....	490
Z. KARIMOVA. RAVISH SO'Z TURKUMINI MA'NOVIY TUZILISHIGA KO'RA GRAMMATIK POS TEGGLASH (PARTS-OF-SPEECH TAGGING) MASALASI	494
SH.R. ABDULLAYEVA. ARTIFICIAL INTELLIGENCE AND ONTOLOGICAL DICTIONARIES: MODERN APPROACHES.....	496
N.I. ADIZOVA. KORPUSGA ASOSLANGAN TURIZMGA OID MATNLARNING O'RGANILISHI	499
U.A. G'OYIBOV. INGLIZ VA O'ZBEK TILLARI FRAZELOGIK BIRLIKLARIDAGI ANTROPONIMIK KOMPLEMENTLARNING O'ZIGA XOS XUSUSIYATLARI	505
M. JURAEV. HISTORICAL FOUNDATIONS AND EARLY FRAMEWORKS OF TEXT COMPLEXITY ASSESSMENT	510
M.N. JO'RAYEVA. O'ZBEK TILIDA KOD ALMASHINUVI	516
N. ABDURAXMONOVA, L. BOBOXO'JAYEVA. TURLI TIZIMLI TILLAR DOIRASIDA ONTOLOGIYANING NAZARIY ASOSLARI	520
N. ABDUJABBAROVA. PARALLEL KORPUSNING O'RNI VA AHAMIYATI.....	524
M. XUDOYBERGANOVA. QORAQALPOQ VA O'ZBEK TILLARINING SINTAKTIK MODELLARINI TEGGLASH MASALASI.....	528
O.O. OBIDJONOVA. TEZAVURUS VA TEZAVURUS LUG'ATLARNI YARATISHGA DOIR	533
S.A. ABDURAXMONOVA. O'ZBEK BOLALAR NUTQINING KORPUSGA ASOSLANGAN PSIXOLINGVISTIK TADQIQI: CHILDES TIZIMI ASOSIDA YONDASHUV	538
M.S. JO'RAQULOVA, S.I. QO'ZIYEVA. O'ZBEK TILIDAGI SO'Z BIRIKMALARINI FORMALLASHTIRISH BOSQICHLARI	541
F. MENGZIYOTOVA. O'ZBEKISTONDA KOMPYUTER LEKSIKOGRAFIYASI YUZASIDAN OLIB BORILGAN TADQIQOTLAR	546
G.I. KUVONDIKOVA. THEORETICAL ISSUES OF CREATING AN EDUCATIONAL CORPUS OF STUDENT SPEECH	549
M. SOBIROVA A. MADADOVA. O'ZBEK TILI MORFOLOGIYASINING ONTOLOGIK MODEL.....	555
I. ASKAROV. DIAKRONIK KORPUSLAR SHAKLLANISH EVOLYUTSIYASI	558
U. ABDURAXMONOV. IKKI TILDAGI FRAZELOGIK BIRIKMALARNING PARALLEL KORPUSDA BERILISHI	562
D.T. GIYASOVA, A.R. ISKANDAROVA. O'ZBEK VA RUS MILLIY KORPUSINING ARHITEKTURASI	566
D.X. XOLBEKOVA. ULUG'BEK HAMDAM IJODIDA LINGVOSTATISTIK TAHLIL	570

A.VOSILJONOV. ELEKTRON MATN KORPUSLARI VA ULARNING DIALEKTOLOGIK TADQIQOTLARDAGI AHAMIYATI	573
F. TOPILDIYEVA, S.A. ERKINOVA. NORMATIV-HUQUQIY HUJJATLARNING HUQUQIY TIZIMDAGI O'RNI	575
F. TOPILDIYEVA, S.A. ERKINOVA. FHDYO HUJJATLARI LOYIHALARINO LINGVISTIK EKSPERTIZADAN O'TKAZISH	577
G.K. XUDAYBERGENOVA. QORAQALPOQ TILINI KORPUS LINGVISTIKASI ASOSIDA TADQIQ QILISH	581
Z. UBAYDULLAYEVA. DUNYO KOMPYUTER LINGVISTIKASIDA SOHAVIY TEZAVURUS YARATISHGA OID TAJRIBALAR TAHLILI	585
I.J. JAHONGIROVA. DIGITAL MEDIA TIL O'RGANISHNING KOMMUNIKATIV KOMPETENSIYASI SIFATIDA	589
R. OMONOV. SUN'IY INTELLEKT YORDAMIDA TIL O'RGANISH: TEXNOLOGIYALAR TA'LIMYIY SAMARADORLIGI VA INSON OMILI MUVOZANATI	594
E.SH. NAZIROVA, F.F. BOYMURODOV. TABIIY TIL MATNLARINI AVTOMATIK GRAMMATIK VA ORFOGRAFIK TAHRIRLASHDA ALGORITMIK YONDASHUVLAR TAHLILI	600
Z.T. TAXIROV. NUTQ VA USLUB	604
A.N. AXMEDOVA. YOZMA TARJIMA KO'NIKMASINI AVTOMATLASHTIRILGAN TARJIMA VA LINGVISTIK TAHLIL VOSITALARI YORDAMIDA OSHIRISH.....	607
D. MIRVALIYEVA. YURIDIK DISKURS MATNLARI.....	610
F.T. ALIJONOVA. O'ZBEKCHA MATNLARNI SEMANTIK TAHLIL QILISH ALGORITMI: MUAMMO VA YECHIMLAR.....	612
N. O'RINBOYEVA. NERGA OID METODOLOGIK YONDOSHUVLAR TAHLILI.....	615
SH.N. MANSUROVA. O'ZBEK TILIDAGI MATNLARDA SIFAT SO'Z TURKUMIGA TEGISHLI SO'ZLARNI POS TEGILASH.....	619
S. RAHMATOVA. MATN QISMLARINING BOG'LANISHIDA TAKRORNING AHAMIYATI.....	622
D.B. DJALILOVA. FRANSUZ VA O'ZBEK MAQOLLARIDA OLAM MANZARASINING AKS ETISHI.....	627

PARATRANSLATOR – PARALLEL KORPUSGA ASOSLANGAN ELEKTRON KONTEKSTOLOGIK TARJIMA LUG‘AT PLATFORMASINI YARATISHNING AYRIM MASALALARI XUSUSIDA

Abduraxmonova N.

DSc, prof., O‘zbekiston Milliy universiteti
Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası mudiri

Tuliyev Ulug‘bek

PhD, O‘zbekiston Milliy universiteti katta o‘qituvchisi

Shamiyeva Gulshoda

O‘zbekiston Milliy universiteti tadqiqotchisi

Annotatsiya: Ushbu tezisdá amaliy loyiha doirasida amalga oshirilgan parallel korpusga asoslangan elektron tarjima lug‘at platformasini yaratishning lingvistik masalalari tahlilga tortilgan. Unga ko‘ra parallel korpus yaratish uzoq jarayonni o‘z ichiga oluvchi murakkab texnologik bosqich sanaladi. Ayniqsa, ko‘p tilli parallel korpus yaratishda tabiiy tilning adabiy tur va uslublarga xosligi, tilning ichki imkoniyatlari, qolaversa, til bilan uning lisoniy va nutqiy hodisalari o‘rtasida farqli jihatlarini avtomatik qayta ishlashda katta ma‘lumotlar bazasini talab qilishi masalalari yoritilgan.

Kalit so‘zlar: parallel korpus, tarjima muqobillari, segmentlash jarayoni, tarjima xotirasi

Abstract: This thesis analyzes the linguistic issues of creating an electronic translation dictionary platform based on a parallel corpus, implemented within the framework of a applied project. According to it, the creation of a parallel housing is a complex technological stage that includes a long process. In particular, when creating a multilingual parallel corpus, the specifics of natural language to literary types and styles, the internal capabilities of the language, as well as the need for a large database for automatic processing of differences between language and its linguistic and speech phenomena are highlighted.

Keywords: parallel corpus, translation alternatives, segmentation process, translation memory

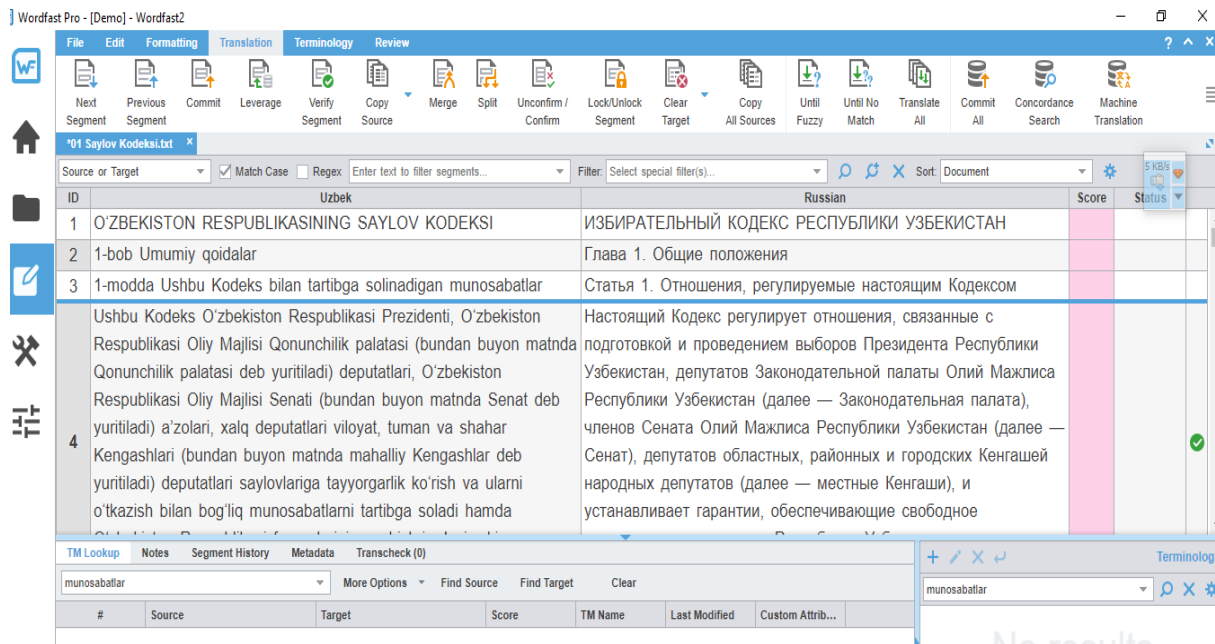
Аннотация: В данной тезис анализируются лингвистические вопросы создания электронной переводческой словарной платформы на основе параллельного корпуса, реализованной в рамках прикладного проекта. Согласно этому, создание параллельного корпуса является сложным технологическим этапом, который включает в себя длительный процесс. В частности, при создании многоязычного параллельного корпуса подчеркивается специфичность естественного языка к литературным типам и стилям, внутренние возможности языка, а также необходимость большой базы данных для автоматической обработки различий между языком и его языковыми и речевыми явлениями.

Ключевые слова: параллельный корпус, альтернативы перевода, процесс сегментации, память перевода

Ushbu tadqiqot IL-7823051524 sonli “PARATRANSLATOR – parallel korpusga asoslangan elektron kontekstologik tarjima lug‘at platformasini yaratish” nomli amaliy loyiha (2024-2025) doirasida amalga oshirildi.

Tarjimaning nechog‘lik sifatli tarjima qilinishi so‘z, so‘z birikmasi yoki iboralarning uslub jihatdan to‘g‘ri tanlanganligi tahlil qilinib, segmentlash yoki tarjima jarayonida yo‘l qo‘yilgan xatolar tarjimon va tilshunoslar tomonidan qayta tahrirlanish jarayoniga ham bog‘liqdir. So‘z

konkordanslarini aniqlash va ilmiy uslubdagi matnda qo‘llangan terminlar uchun maxsus dastur yordamida ma’lumotlar bazasining shakllantirilishi, ajratib olingan parallel matnlardagi til va nutqiy birliklar alohida jadval shaklga keltirilishi tarjima dasturi uchun dastlabki bosqichlardan biridir. Segmentlash jarayoni to‘liq yakuniga yetkanidan so‘ng barcha parallel matnlar tarjima xotirasiga yuboriladi (1-rasm).



1-rasm

Shuningdek, ajratilgan gap segmentlaridan terminlar bazasi, so‘z va so‘z birikmalari avtomatik ravishda aniqlanib, inson tomonidan uning muqobillik darajasi tahrir qilinadi.

Parallel korpusdan ma’lumotlar bazasi sifatida jamlangan tarjima xotirasi muhitidan foydalanish inson tomonidan qilinadigan tarjima jarayonini tezlashtiribgina qolmay, uning sifatiga ham ijobiy ta’sir ko‘rsatadi. Zero matnda yo‘l qo‘yilgan xatolarni ko‘rsatish va to‘g‘irlash, qolaversa, turli uslubga tegishli matnlardan ma’nodosh so‘zlarning mosini qo‘yishda ushbu dasturning imkoniyatlarini qo‘llash mumkin. Parallel matnlar uchun segmentlash va lingvistik tahlil juda muhim bosqichlardan sanaladi. Parallel matnlardagi konkordanslarni aniqlashda *so‘z, so‘z birikmasi yoki barqaror birikmalarning* u yoki bu tildagi muqobilligiga asoslaniladi. Bu holatda adekvat tarjima doim ham o‘zini oqlamaydi, chunki ayrim so‘zlar tushib qolishi yoki biror komponent qo‘shilishi hisobiga ularning leksik va grammatik modeliga o‘zgartirish kiritiladi. Shu bois parallel matnlardagi kalit so‘zlar uchun kontekstda tez-tez uchrab, tarjimasida muayyan o‘zgarishga uchramagan faol so‘z va terminlar ma’lumotlar bazasida yig‘iladi. Ushbu holatda o‘zbek tilidagi yordamchi so‘z turkumlari, mustaqil ma’no anglatmaydigan so‘zlar (taqlid, undov, ko‘makchi fe‘llar, nomustaqil fe‘l va h.k.) hamda matnda omonim sifatida ko‘p qo‘llanadigan so‘zlarni ushbu MBdan holi etiladi (2-rasm).

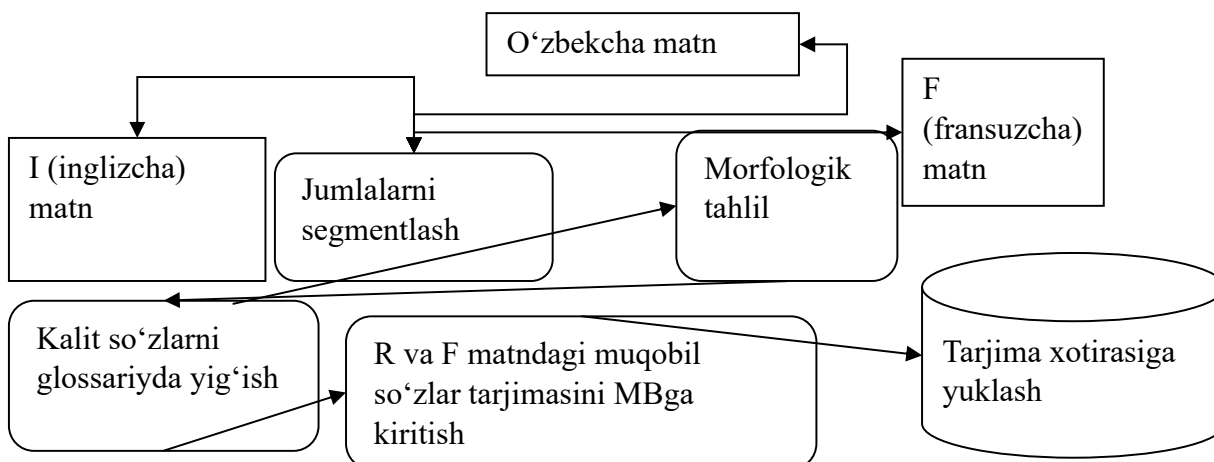
ID	Uzbek_term	Russian_term
1	SAYLOV KODEKSI	ИЗБИРАТЕЛЬНЫЙ КОДЕКС
2	O'ZBEKISTON RESPUBLIKASINING SAYLOV KODEKSI	ИЗБИРАТЕЛЬНЫЙ КОДЕКС РЕСПУБЛИКИ
3	munosabatlar	отношения
4	kodeks	кодекс
5	O'zbekiston Respublikasi Prezidenti	Президента Республики Узбекистан
6	Oliy Majlisi Qonunchilik palatasi deputatlari	депутатов Законодательной палаты

2-rasm

Chunki tarjimada ularning ma’nosini oldindan bashorat qilishning imkoni mavjud emas. O‘zbek tilida qo‘shma so‘zlarning miqdori nisbatan ko‘p bo‘lib, quyidagicha bo‘lishi mumkin: $\{n...S / S...n1\} \Rightarrow$ qo‘shma so‘z (*nuqtayi nazar / nazar qilmaslik*)

$$\{n...SW_n\} \Rightarrow \text{ibora (nazari past bo'lmoq)}$$

Bizning parallel korpusimiz ilmiy, rasmiy, badiiy, ommabop uslubga tegishli matnlarni o‘z ichiga oladi. Parallel matnlarni tahlil qilishning algoritmi quyidagicha:



CAT muhitida tarjima xotirasiga yuklangan parallel korpuslar tarjimon va lingvistlar tomonidan qayta tekshirilishi lozim. Yuqorida aks etgan algoritmnning morfologik tahlil bosqichi muhim bosqich sanaladi. Chunki o'zbek tilining agglyutinativ xususiyati matnda token sifatida qatnashuvchi so'zshakllarda namoyon bo'ladi, bu esa ruscha yoki inglizcha so'zlarning muqobil konkordansini berishda juda muhim ahamiyat kasb etadi. Shu bois o'zbek tilidagi token va lemmalar ajratilgandan so'ng ular mutaxassislar tomonidan juftliklar bo'yicha bazaga kiritiladi.

Ilmiy matnlardagi doimiy ishlatiladigan me'yoriy qoliplar sifatida quyidagi kalit soʻzlarga e'tibor beriladi: soʻz *asosi*=> *derivatsion shakli*=> *soʻz shakli* => *koʻp qoʻllanilgan soʻz, fraza yoki terminlarning chastotasi*=> *konkordansi*=> *tarjimasi*. Namunada koʻrsatilgan matn “uzun” jumlati matn turiga kiradi, chunki unda tinish belgilar, qolaversa, predikativ elementlar miqdori koʻp boʻlib, ushbu matn tarjimon tomonidan ikki qismga ajratilib tarjima qilingan. Yirik matnli matndan ajratib olingan parallel jumlarlar generatsiyasi quyidagi bosqichlarda amalga oshirilishi mumkin ¹:

- I. Parallel matndan olingan tokenlashgan soʻzning muqobil juftlari haqida maʼlumot olinadi;
- II. Uzun jumlati matn muayyan tinish belgilari “,”, “;”, “:” boʻlgan qismigacha alohida segment gaplarga ajratiladi;
- III. Asos – tarjima qilingan gaplarning segmentlari tarkibining miqdori aniqlanib, muvofiqlik holati belgilanadi;
- IV. Tarjima – asos matnlarning muqobillik holati aniqlanadi;

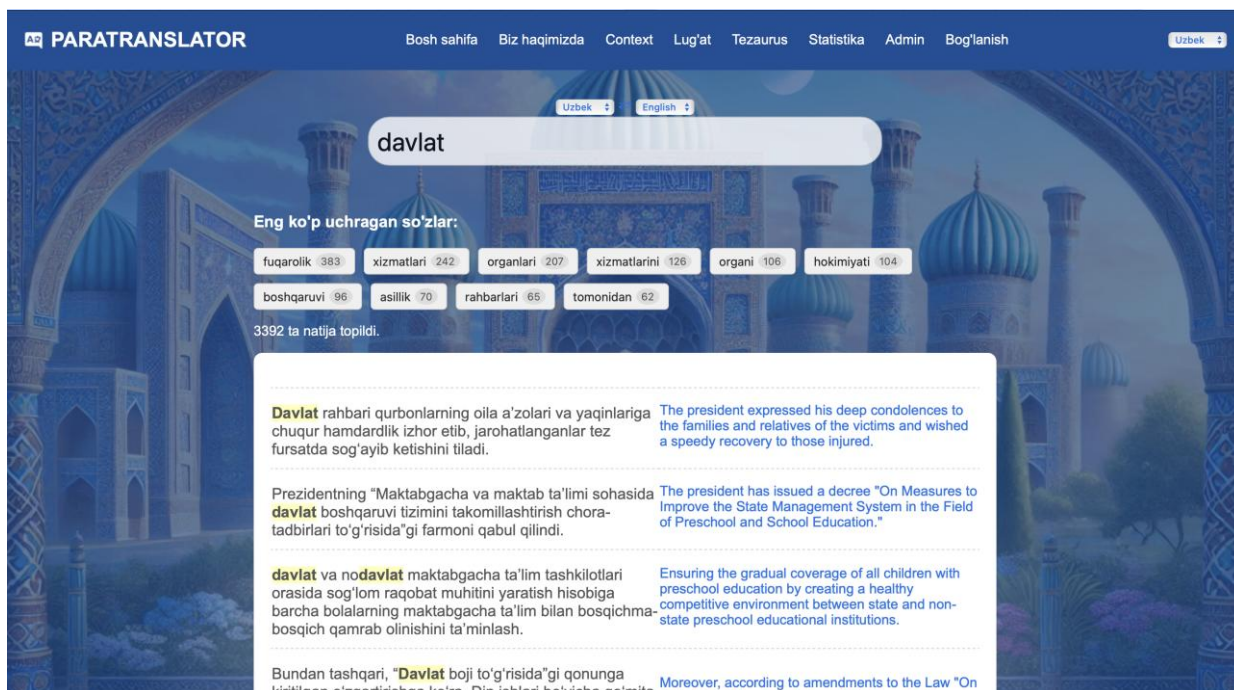
Segmentlar o'rtasida birga ko'p yoki ko'pga ko'p munosabati kuzatilganda birga bir munosabat shaklida bir nechta segment birliklar o'zaro bog'lanadi yoki birlashtiriladi.

Quyidagi 3-rasmda “PARATRANSLATOR” – parallel korpusga asoslangan elektron kontekstologik tarjima lug‘at platformasini yaratish” nomli amaliy loyihamiz doirasida yaratilgan saytida parallel korpus interfeysi berilgan: <https://paratranslator.uz/>

Konkordanslarning segmentlashuvi mashina tarjimasida ikkinchi tildagi soʻz birikmasi yoki muayyan turgʻun birikmaning sinonimlarini toʻgʻri tanlashda xizmat qiladi. Shu bois ular ham tarjima xotirasiga lingvistik baza sifatida eksport qilinadi. Natijada asos til *a-seg_a* va tarjima qilingan til *t-seg_i* oʻrtasidagi segment birliklarning (*jumla, soʻz birikmasi, fraza, termin*) muqobillik bazasi shakllantiriladi. Shunday qilib, jumlarlar segmentlashuvidan parallel korpus muqobilligi yaratiladi. Oʻzbek tilida ikki xil lotin va kirill grafemalarning amalda qoʻllanishi ham

¹ Коптев М. Введение в корпусную лингвистику. – Прага, 2014. – С.129-132.

morfologik tahlil jarayonida o'zining ta'sirini ko'rsatadi. Negaki, lingvistik baza har ikki grafema uchun tayyorlanishini taqozo etadi.



3-rasm

Tarjimon tomonidan amalga oshiriladigan tarjima o'zgarishlari (translation shift)ni matnlar o'rtasidagi *sintaktik*, *morfologik*, *leksik* va *semantik* jihatdan kuzatish mumkin. Parallel matnda shunday holatlarni erkin kuzatish, kontekstdan kelib chiqib, so'zlarning o'zaro muqobillarini aniqlash imkoniyati tug'iladi. Kalit so'zlar esa so'zning nechog'lik kontekstda qo'llanganlik chastotasiga qarab belgilanadi.

Parallel korpus nafaqat mashina tarjimasi uchun lingvistik resurs, balki bilingval lug'atshunoslik, qiyosiy tarjima, chog'ishtirma tilshunoslik kabi sohalar uchun muhim manba bo'lib hizmat qiladi. Endilikda neyro yoki statistik mashina tarjimasi uchun parallel matnlarning ma'lumotlar bazasini yaratish va uning lingvistik jihatlarini chuqur o'rganishning asosiy obyekti sanaladi. Nechog'lik tarjima xotirasiga yuklangan tarjima matnlarning muqobil juftliklari aniqlanib, boyitib borilsa, kelajakda turli uslubga tegishli avtomatik tarjima qiluvchi tarjimon dasturlarning sifatini yaxshilashda qo'l keladi. Shuningdek, parallel ravishda yaratilgan og'zaki (Parallel WaveNet) va yozma matnlar nutq sintezatoriga doir tadqiqotlar olib borishda mashina fondi vazifasini o'tamoqda².

“PARATRANSLATOR” – parallel korpusga asoslangan elektron kontekstologik tarjima lug'at platformasini yaratish” nomli amaliy loyiha doirasida ingliz, ransuz, rus, ispan, turk, qoraqolpoq, koreys, yapon, xitoy kabi rasmiy, ilmiy, badiiy, publististik uslubdagi matnlar segmentlandi.

Parallel korpusdagi tarjimaning sifatiga bog'liq ba'zi muammolar mavjud bo'lib, ularni bartaraf etish mashina tarjimasining sifatiga ham ta'sir o'tkazadi. Quyida shu kabi bir nechta muammolar tahlilini keltirib o'tamiz:

1. Parallel korpusdagi tarjimaning sifatiga bog'liq muammolar. Parallel korpuslardan foydalanishda duch keladigan eng katta muammolardan biri — *tarjimaning sifatidir*. Korpusdagi tarjimalar har doim ham to'g'ri yoki aniq bo'lmasligi mumkin, chunki tarjima qiluvchilar orasida stilistik, grammatik yoki semantik xatolar bo'lishi ehtimoli mavjud. Bu, ayniqsa, kam resursli tillarda ko'p uchraydi. Korpusdagi tarjimalarning sifatini ta'minlashda *noaniqliklar* va *polisemik*

² Abduraxmonova N. Korpus lingvistikasi (darslik). – Germaniya: Globeedit 2023. –B.357.

(ko'p ma'noli) so'zlar, shuningdek, so'z birikmalarining kontekstga mos kelmasligi, tarjima tizimining sifatini pasaytirishi mumkin³. Buning oldini olish uchun bir nechta usullar taklif etiladi. Masalan,

1. *Tarjima sifatini tahrirlash*: Tarjimalarni qo'lda tahrir qilish orqali xatolarni aniqlash va tuzatish mumkin. Ushbu jarayon, parallel korpuslarni tayyorlashda sifatni yaxshilash uchun muhim vosita hisoblanadi.

2. *Yuqori sifatli parallel korpuslar yaratish*: Korpusni yaratishda manbalarni aniq tanlash va tarjima jarayonini qat'iy nazorat qilish kerak. Masalan, yirik va ishonchli parallel korpuslardan foydalanish (masalan, Europarl, OpenSubtitles) yuqori sifatli tarjimalarni taqdim etadi⁴.

3. *Avtomatik baholash*: Tarjimaning sifatini baholashda BLEU (*Bilingual Evaluation Understudy*) kabi metrikalardan foydalanish orqali tarjima sifatini baholash va yanada optimallashtirish mumkin.

2. Korpusning hajmiga bog'liq bo'lgan muammolar. Parallel korpuslarning samarali ishlashi uchun ularning hajmi katta bo'lishi talab etiladi. Korpusdagi matnlar sonining kamligi yoki o'rta hajmdagi ma'lumotlar sifatli tarjima uchun yetarli bo'lmisligi bilan birga, kichik hajmdagi parallel korpuslar bilan ishlashda modelning o'rganish imkoniyatlari cheklangan bo'ladi, bu esa tarjima tizimining samaradorligini pasaytirishi mumkin. Shu bilan birga, ba'zi tillarda parallel korpuslar yetarli darajada mavjud emas, bu esa kam resursli tillarda mashina tarjimasi tizimlarini o'rganishda qiyinchiliklarga olib keladi⁵. Buning oldini olish uchun esa quyidagi usullar samarali hisoblanadi:

1. *Korpusni kengaytirish*: Ba'zi tillarda parallel korpuslar yetarli bo'lmagan hollarda, korpusni kengaytirish usullari, masalan, tarjima xizmati yoki foydalanuvchi generatsiyasi orqali parallel ma'lumotlarni yaratish mumkin. Shuningdek, monolingual korpuslardan (bir tilda bo'lgan matnlar) yarim avtomatik tarzda parallel korpuslar yaratish usullari ham qo'llanilishi mumkin.

2. *O'zaro ko'p tilli o'rganish*: Ba'zi hollarda til resurslari cheklangan tillarda, o'rganish jarayonida ko'p tilli o'rganish yondashuvi ishlatiladi. Bu orqali ba'zi tillarni o'rgatishda boshqa tillarda o'rganishdan foydalanish mumkin⁶.

3. Korpuslarni sinxronlashtirishdagi muammolar. Parallel korpuslarda matnlarning sinxronlashtirilishi (so'zlar yoki iboralarining mos kelishi) va tarjima birikmalarining to'g'ri joylashishi katta ahamiyatga ega. Sinxronizatsiya qilishda yuzaga keladigan xatoliklar tarjima tizimining to'g'ri ishlashiga to'sqinlik qilishi mumkin. Korpusdagi ba'zi iboralar yoki jumlar o'rtasidagi bog'lanishlar noto'g'ri bo'lishi esa tarjima sifatiga salbiy ta'sir ko'rsatadi⁷. Bunday muammolarni bartaraf etish uchun quyidagi usullardan foydalanish mumkin:

1. *Tarjima frazalarini aniqlash va tahlil qilish*: Sinxronizatsiya jarayonini optimallashtirish uchun so'z birikmalarini va frazalarni aniq ajratish parallel korpuslarni to'g'ri ishlashiga yordam beradi.

2. *Avtomatik sinxronizatsiya algoritmlari*: Korpuslarni sinxronlashtirish jarayonida, statistik va neuronal tarmoq asosidagi avtomatik sinxronizatsiya algoritmlaridan foydalanish matnlarning to'g'ri joylashishini va tarjima o'zgartirishlarini optimallashtiradi.

4. Semantik farqlarni e'tiborga olmaslik. Parallel korpuslar, til o'rtasidagi semantik farqlarni to'liq hisobga olmaydi. Ba'zi tillarda aynan bir so'zning ko'plab ma'no variantlari bo'lishi mumkin, bu esa tarjimaning aniqligini pasaytirishi mumkin. Shu sababli ham korpusdagi semantik

³ J. Tiedemann. Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. European Language Resources Association (ELRA).

⁴ P.Koehn. Europarl: A Parallel Corpus for Statistical Machine Translation. *Proceedings of the 10th Machine Translation Summit*, 2005. –P.79-86.

⁵ P.Koehn. Statistical Machine Translation. *Cambridge University Press*. 2010. – p.487

⁶ B.Zoph, K.Knighk. Multi-Source Neural Machine Translation. *Proceedings of the NAACL-HLT 2016*, –P.122-132.

⁷ P. Koehn. Europarl: A Parallel Corpus for Statistical Machine Translation. *Proceedings of the 10th Machine Translation Summit*, 2005. –P.79-86.

mosliklar va farqlarni o'rganish jarayonida ba'zan so'zlar va iboralar o'rtasidagi kontekstual o'zgarishlar noto'g'ri tarjima qilishga olib keladi⁸.

Parallel korpuslar, mashina tarjimasi va tabiiy tilni qayta ishlash sohasida katta ahamiyatga ega bo'lsa-da, ularni samarali ishlatishda ba'zi muammolar mavjud. Tarjimaning sifatiga bog'liq muammolar, korpusning hajmi, sinxronizatsiya muammolari, semantik farqlar va kichik tillar kabi cheklovlar bu jarayonni murakkablashtiradi. Biroq, zamonaviy yondashuvlar, masalan, avtomatik sinxronizatsiya algoritmlari, attention mexanizmlari va kam resursli tillarda o'rganish kabi texnikalar ushbu cheklovlarni yengib o'tish va parallel korpuslardan samarali foydalanish imkonini beradi.

Foydalanilgan adabiyotlar:

1. J. Tiedemann. Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. European Language Resources Association (ELRA).
2. P.Koehn. Europarl: A Parallel Corpus for Statistical Machine Translation. *Proceedings of the 10th Machine Translation Summit*, 2005. –P.79-86.
¹P.Koehn. Statistical Machine Translation. *Cambridge University Press*. 2010. – p.487
3. B.Zoph, K.Knighk. Multi-Source Neural Machine Translation. *Proceedings of the NAACL-HLT 2016*, –P.122-132.
4. Abduraxmonova N. Korpus lingvistikasi (darslik). – Germaniya: Globeedit 2023. –B.357.
5. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzworonet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
6. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzworonet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).

⁸ A. Vaswani, N. Shazee, N. Parma, J. Uszkorei, L. Jones, N. A. Gomez, L. Kaiser, I. Polosukhin. *Attention is All You Need*. *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)*, 30, 5998-6008.

THE USE OF WEBQUESTS IN FOREIGN LANGUAGE TEACHING TO FOSTER STUDENTS' COMMUNICATIVE COMPETENCE

Zhuparbek D.

L.N. Gumilyov Eurasian National University

Nurkenova S.S.

L.N. Gumilyov Eurasian National University

Аннотация: Современное образование активно внедряет инновационные технологии, среди которых особый интерес представляют веб-квесты. Данная работа посвящена изучению эффективности использования веб-квестов в качестве инновационного инструмента для развития коммуникативных навыков учеников, обучающихся иностранным языкам. В исследовании рассматривается, как веб-квесты могут способствовать повышению мотивации учащихся, развитию их критического мышления, сотрудничества и умения работать с информацией. Цель исследования: определить влияние веб-квестов на формирование коммуникативной компетенции учеников, обучающихся иностранным языкам. Объектом исследования является процесс развития коммуникативных навыков учащихся. Предметом исследования является технология веб-квеста как способ развития коммуникативных навыков учеников средней степени обучения.

Ключевые слова: веб-квест, поисковая деятельность, интернет-технологии, коммуникативная компетенция.

Abstract: Modern education actively introduces innovative technologies, among which web quests are of particular interest. This work is devoted to studying the effectiveness of using web quests as an innovative tool for developing the communicative skills of students learning foreign languages. The study examines how web quests can contribute to increasing students' motivation, developing their critical thinking, cooperation, and ability to work with information. The purpose of the study is to determine the impact of web quests on the formation of communicative competence in students learning foreign languages. The object of the study is the process of developing students' communicative skills. The subject of the study is web quest technology as a way to develop the communicative skills of middle school students.

Keywords: web quest, search activity, internet technologies, communicative competence.

Introduction

The rapid advancement of information and communication technologies is transforming all spheres of human activity, including education. Digital tools offer new opportunities to enhance learning outcomes and foster a wide range of learner competencies. One such tool is the webquest—an interactive task that promotes autonomous exploration of a topic through guided internet research. Unlike traditional methods of knowledge transmission, webquests engage students in active learning, encouraging them to search for, analyze, and synthesize information. This process develops critical thinking, information literacy, collaboration, and communication skills.

The relevance of this study lies in the increasing integration of digital technologies into education and the growing need to develop learners' communicative competence in a globalized world. In particular, the use of online testing in foreign language education demonstrates considerable potential for improving instructional effectiveness and supporting the development of independent learning skills. Contemporary pedagogical approaches emphasize learner autonomy and lifelong learning, necessitating a shift from knowledge transmission to the cultivation of cognitive and metacognitive abilities. This shift calls for a reassessment of teaching

methodologies and the modernization of both pedagogical practices and the technological infrastructure of the learning environment.

Literature Review

A well-established definition of the webquest in the context of foreign language education is provided by T.A. Goltsova and E.A. Protsenko, who describe it as “a specially organized research-based project grounded in the principles of problem-based and inquiry learning, integrating a range of methods and techniques that foster the development of information-analytical skills and implemented primarily through internet resources” [2].

Research by scholars such as D.V. Larionova highlights the motivational potential of educational webquests, emphasizing their role in increasing student engagement [3]. E.A. Ryabukha and E.A. Gafarova further characterize webquests as tools for enhancing learners' sociocultural competence in a foreign language context [4]. These perspectives converge on the idea that the development of communicative competence is a priority in contemporary language education, particularly at the tertiary level.

Webquests offer several pedagogical advantages that contribute to more effective and student-centered learning. According to T.N. Dimcheva, they promote learner engagement, critical thinking, and information literacy [5]. Scholars such as A.L. Portilla and V.G. Camacho explore future directions in webquest development, including the integration of mobile technologies, social media, and virtual reality [6]. Nonetheless, challenges remain—such as the need for ongoing teacher training and the risk of cognitive overload. Educators must stay informed about technological advancements and address these issues to optimize the use of webquests in instructional settings.

Global research also confirms the effectiveness of webquests in developing language skills. For instance, a study by S. Ebadi and M. Rahimi found a significant improvement in students' English reading skills following the implementation of webquests [7]. By offering opportunities for authentic language use, vocabulary expansion, and contextualized reading tasks, webquests support the development of both receptive and productive skills. Their interactive nature encourages learners to communicate, collaborate, and present information in English, thus fostering speaking and writing competencies.

N.J. Mongyi emphasizes the role of webquests in promoting active learning and learner autonomy, enabling students to navigate online resources, make informed decisions, and engage in problem-solving independently [8]. Through inquiry-based tasks, learners enhance their critical thinking, digital literacy, and research skills.

An additional benefit of webquests is their ability to incorporate authentic and relevant online materials into classroom activities without requiring excessive time for information retrieval. The resources (links) used in webquests are curated by the teacher in advance, ensuring that the content is tailored in both quality and quantity to the needs of the target group (e.g., a specific course or seminar). Importantly, the teacher designs the webquest, selecting freely available internet content and integrating it into the lesson structure. Supplementary materials such as textbooks or handouts can also be provided during in-person sessions, allowing the learning experience to remain interactive and resource-rich.

Table 1. Comparison of Learning Stages

Bloom's Taxonomy	Development of communicative competence	Stages of WebQuest Methodology (Dodge & March)	Description
Knowledge	Passive knowledge of vocabulary and grammar	Introduction	Students are introduced to the topic; background information is provided, and interest is generated. This stage corresponds to recalling basic linguistic elements.

Comprehension	Understanding the meaning of spoken or written language	Task	The central task is presented. Students comprehend the communicative goals and how the language is used within a meaningful context.
Application	Use of language in basic communicative situations	Process	Students follow guided steps to complete the task, applying language knowledge in structured, real-life scenarios.
Analysis	Analysis of language structures; comparison of linguistic features	Resources	Students work with selected authentic materials to analyze, extract, and interpret language patterns and usage.
Synthesis	Producing original utterances and texts	Product	Students create their own spoken or written content (e.g., presentations, dialogues, essays), integrating language knowledge with creative expression.
Evaluation	Evaluating the correctness and effectiveness of one's own and others' language use	Conclusion	Students reflect on their performance, assess the quality of their language production, and provide or receive peer feedback.

The methodology of web quests, developed by Dodge March, and Bloom's Taxonomy are two powerful tools that, when combined, enhance the educational process and foster the development of a wide range of skills in students, including communicative skills. A web quest structures the learning activity by presenting students with a series of tasks that they must solve using various sources of information. Bloom's Taxonomy, on the other hand, classifies cognitive processes across different levels of complexity, from remembering to creating. The synthesis of these two tools allows the creation of learning scenarios that not only develop knowledge and skills but also foster communicative competence.

In this study, we explore why and how teachers integrate web quests into their English language teaching curriculum. Our research consisted of a qualitative investigation with two parts. In the first part, we examined the activities of 25 trainee teachers. These teachers developed web quests and implemented them in their lessons. The results of this phase revealed that the majority of teachers are willing to use web quests in their future lessons.

In the second part, teachers who had already used web quests in their classrooms were surveyed with six questions about their experiences (the results are presented as percentages).

Table 2. Results of Teacher Surveys

№	Questions	Survey results
«1	Do you use webquests in your foreign language teaching practice?	a) Yes, regularly (once a month or more often) – 30% b) Yes, sometimes (once every few months) – 45% c) Yes, but very rarely (once every six months) – 25%
«2	What aspects of communicative competence, in your opinion, are most effectively developed in students using web quests? (participants could choose several options)	a) Speaking skills – 68% b) Listening skills – 55% c) Reading skills – 42% d) Writing skills – 37% e) Teamwork – 60%
i3	What types of webquests do you find most useful for improving communication skills?	a) Search web quests – 35% b) Creative web quests – 40% c) Role-playing web quests – 55% d) Research web quests – 25%

4	What difficulties do you encounter when developing or implementing webquests in the educational process? (participants could select several options)	a) Technical difficulties – 40% b) Lack of time for preparation – 60% c) Lack of teaching materials – 35% d) Low student engagement – 20% e) Lack of digital skills among teachers – 30%
5	How do you assess the impact of webquests on students' motivation to learn a foreign language?	a) Significantly increases motivation – 45% b) Moderately increases motivation – 40% c) No effect noticed – 10% d) Motivation decreases – 5%
6	What additional methods or technologies do you use along with web quests to improve students' communicative competence?	a) Using interactive platforms (e.g. Quizlet, Kahoot) – 65% b) Group projects and discussions – 50% c) Video and audio materials – 55% d) Games as a teaching method – 35%

The data obtained clearly demonstrate how web quests are utilized in education and their effectiveness in developing communicative skills. Teachers noted that web quests are more effective than previously employed methods. Additionally, recommendations for creating and utilizing web quests, proposed by experienced educators, are presented. These recommendations can be summarized as follows:

- Plan the web quest carefully and thoughtfully.
- Be patient and flexible.
- Set challenging yet achievable tasks for students.
- Clearly explain the process and provide ongoing support.
- Select high-quality resources and ensure they are kept up-to-date.
- Teach students to collaborate effectively.
- Support students in their collaborative efforts.
- Prepare students for solving technical problems and equip them with necessary computer skills.

Conclusion

In conclusion, it can be noted that web quest-based learning strategies address the limitations of traditional teaching methods, fostering a more flexible and effective collaborative learning environment. At the core of a web quest are research tasks that help novice learners master the learning process. Teachers monitor students' work, assess their progress, and encourage them to reflect on their achievements after completing the web quest. Teachers also report that peer assessment and feedback helped reduce students' anxiety regarding evaluation. Web quests are a foundation of constructivist learning and, as demonstrated, effectively develop communicative skills while positively influencing the development of "soft skills" that are crucial for both education and employment.

It is also recommended that educators pay particular attention to elements such as hints, active learning, deep learning, peer collaboration, and encouraging students to take responsibility for their own learning. The use of web quests in foreign language lessons can enhance communicative competence through the following factors:

- Authenticity of communication: Web tasks replicate professional and everyday situations, creating realistic communication scenarios.
- Teamwork: Students can collaborate to solve problems, exchange information, and find common solutions.
- Development of information search and processing skills: Web tasks encourage students to actively search for, analyze, and integrate information from various sources.
- Development of presentation skills: Students learn to prepare presentations, present research findings, defend their viewpoints, and engage effectively with an audience.

Thus, educational web quests represent an effective tool for developing students' communicative skills, enabling them to successfully adapt to the demands of the modern

workplace. This study also corroborates findings from other research, which suggest that time constraints are one of the barriers to implementing web-based assessments. Teachers further reported that they require both visual and verbal instructions to facilitate the understanding of web tasks. Another issue raised in this study concerned communication during task completion: it was found that most web-based tasks do not focus on communication, which does not motivate elementary school students to form communicative groups. Finally, students require additional guidance from teachers.

References:

1. Сапух Т.В. Возможности технологии «веб-квест» для развития иноязычной коммуникативной компетенции // Научно-педагогическое обозрение. Pedagogical review. – 2017. – 4(18). – 149-157.
2. Гольцова Т.А., Проценко Е.А. Применение веб-квест технологии при обучении межкультурной коммуникации // Отечественная и зарубежная педагогика. – 2021. – 1(2). – 89-96.
3. Ларионова Д.В. Технологии веб-квеста в создании педагогами медиаобразовательной среды // Мир науки, культуры, образования. – 2020. – 1(80). – 30-32.
4. Рябуха Е.А., Гафарова Е.А. Веб-квесты, как форма предъявления творческих задач в формате дистанционного обучения // Информационные технологии как основа эффективного инновационного развития. – 2021. – 4(6). – 215-219.
5. Димчева Т.Н. Актуальность применения веб-квестов в обучении иностранному языку общения и педагогические основы их проектирования // Молодой исследователь: вызовы, поиски и перспективы развития российского образования. – 2017. – 42-48.
6. Portilla A.L., Camacho V.G. A webquest tool to develop communicative competence in EFL students with an A2 proficiency level // Zona Próxima. – 2017. – 26. – 82-98.
7. Ebadi S., Rahimi M. An exploration into the impact of WebQuest-based classroom on EFL learners' critical thinking and academic writing skills: A mixed-methods study // Computer Assisted Language Learning. – 2018. – 5-6. – 617-651.
8. Mongyi N.J. Developing learners' intercultural communicative competence using thinking routines in Web Quests English language teaching // Mester és Tanítvány. – 2024. – 2(1). – 33-46.

A CORPUS-BASED DESCRIPTION OF GRAMMATICAL ERRORS IN UZBEK EFL LEARNER WRITING

Amirsheibani Morteza

Faculty of Languages and Preschool Education,
Tashkent Institute of Economics and Pedagogy (TIPI)

Yo'ldoshev Otabek Abdug'aniyevich

Faculty of Languages and Preschool Education,
Tashkent Institute of Economics and Pedagogy (TIPI)

Abstract: This study provides a descriptive analysis of grammatical errors in the written English of Uzbek upper-intermediate (B2 level) learners, using a small-scale learner corpus compiled from authentic student essays in academic lyceums and language schools in Uzbekistan. Drawing on principles of corpus linguistics and second language acquisition, the research aims to identify recurrent error types, categorize them, and uncover patterns of learner interlanguage shaped by L1 influence and developmental factors. The corpus, consisting of approximately 20,000 words, was analyzed using AntConc software to extract frequency data and concordance lines. The most common grammatical errors were found in the use of verb tense/aspect, article usage, subject-verb agreement, and prepositions. Many of these errors appear to stem from negative transfer from Uzbek and Russian, which lack certain grammatical categories found in English. The findings provide insight into the grammatical challenges faced by Uzbek learners and offer implications for grammar instruction, curriculum design, and teacher training in EFL contexts across Central Asia.

Keywords: learner corpus, grammatical errors, Uzbekistan, EFL writing, interlanguage, L1 transfer, corpus linguistics, B2 level learners

Introduction

Over the past few decades, grammar accuracy has remained a central concern in English as a Foreign Language (EFL) instruction, particularly in contexts where English is taught as a school subject rather than used as a medium of communication. In Uzbekistan, a country where English holds growing importance in education and international mobility, the demand for improved learner proficiency has increased significantly. Despite ongoing reforms in curriculum and teacher training, learners often continue to struggle with core areas of English grammar. This is especially evident in their writing, where persistent errors may indicate deeper issues in language acquisition, cross-linguistic influence, and instructional practices (Abduazizov & Umarova, 2022).

Corpus linguistics offers a powerful approach to understanding these issues by enabling researchers to analyze authentic learner data systematically. Learner corpora—collections of written or spoken texts produced by language learners—can reveal patterns of grammatical development, fossilization, and error (Granger, Gilquin, & Meunier, 2015). Unlike intuition-based approaches, corpus-based studies allow for the empirical observation of learner interlanguage—the evolving language system that reflects a learner's developmental stage and L1 background (Selinker, 1972; Ellis, 2008).

In Uzbekistan, most learners are native speakers of Uzbek or Russian, two languages that differ significantly from English in their grammatical systems. For instance, Uzbek is an agglutinative, article-less language with a flexible word order and limited tense distinctions, while Russian, although inflected and tense-marked, also lacks articles and employs different syntactic structures. These typological differences often result in recurring error patterns in learners' English output, particularly in areas like article use, prepositions, subject-verb agreement, and verb tense/aspect (Muminov, 2021; Nazarova & Kadyrova, 2023).

However, there remains a shortage of detailed corpus-based studies focused specifically on grammatical errors among Uzbek EFL learners. While large international learner corpora such as ICLE or EFCAMDAT offer valuable benchmarks, they rarely include learners from Central Asian countries, making it difficult to generalize findings. As a result, local data is essential to understand the specific needs of Uzbek learners and to inform context-sensitive teaching strategies (Hasanova, 2020).

This study aims to fill this gap by compiling and analyzing a small-scale corpus of written texts produced by B2-level Uzbek learners. The goal is to provide a descriptive overview of the most frequent grammatical errors and to explore the possible linguistic and pedagogical reasons behind them. In doing so, the study hopes to contribute to the growing field of learner corpus research and to offer insights that can support more effective grammar instruction in Uzbekistan and similar EFL contexts.

Review of the Literature

Grammatical accuracy in EFL writing has been the focus of extensive research in second language acquisition (SLA), particularly in relation to learner interlanguage development and error analysis. Early work by Corder (1967) emphasized the pedagogical value of learner errors, arguing that errors are not simply mistakes to be corrected, but evidence of the language learning process itself. Later, Selinker (1972) introduced the concept of interlanguage, highlighting how learners develop an intermediate linguistic system influenced by both their native language and the target language. This concept remains foundational in contemporary corpus-based studies of learner language (Ellis, 2008; Granger et al., 2015).

Over the past two decades, the emergence of learner corpora has significantly advanced the study of grammatical errors. Projects such as the International Corpus of Learner English (ICLE) and the Cambridge Learner Corpus (CLC) have enabled researchers to examine L2 writing from multiple L1 backgrounds, identifying both universal developmental patterns and L1-specific influences (Granger et al., 2015; Nesselhauf, 2004). These studies have consistently shown that certain grammatical categories—articles, verb tense/aspect, prepositions, and subject-verb agreement—are particularly error-prone across learner populations.

However, L1 transfer plays a major role in shaping error patterns. According to Odlin (1989), the degree of linguistic distance between a learner's first language and English can significantly affect the type and frequency of errors. For instance, L2 learners whose L1 lacks articles (e.g., Chinese, Russian, Uzbek) tend to omit or misuse articles more frequently than those whose L1 includes them (Ionin, Ko & Wexler, 2004). Similarly, learners from languages with flexible word order or agglutinative structures, such as Uzbek, may show difficulties in maintaining syntactic patterns typical of English (Muminov, 2021).

While global learner corpora have provided broad insights, the underrepresentation of learners from Central Asia—especially Uzbekistan—has limited the applicability of such findings to this region. Research from within the Uzbek context has begun to emerge in recent years. For example, Hasanova (2020) analyzed student writing at tertiary institutions and found frequent misuse of tense forms and prepositions, attributing many errors to negative transfer and insufficient grammar instruction. Similarly, Nazarova and Kadyrova (2023) examined upper-intermediate learners' essays and identified patterns of confusion between simple past and present perfect, a well-documented issue among EFL learners from non-Indo-European language backgrounds.

Despite these efforts, there remains a lack of systematic, corpus-based studies of grammatical errors specific to Uzbek learners. Much of the existing research relies on intuition, classroom observation, or small error samples without rigorous frequency analysis. Given the increasing importance of English proficiency for academic and professional advancement in Uzbekistan, a clearer understanding of learner grammar through corpus analysis is both timely and necessary. As Römer (2022) notes, learner corpora are “an indispensable tool for understanding real learner language use and improving language pedagogy.”

This study builds on previous work by adopting a descriptive, data-driven approach to grammar errors in Uzbek learner writing. By compiling and analyzing a dedicated learner corpus, it seeks to provide empirical evidence of error frequency, examine the influence of the L1, and suggest practical recommendations for improving grammar instruction in Uzbekistan's EFL classrooms.

Methodology

This study employs a descriptive, corpus-based approach to identify and analyze frequent grammatical errors in the written English of Uzbek EFL learners. The research is situated in the context of academic lyceums and private language schools in Tashkent and Samarkand, where English is taught as a compulsory subject at the upper-intermediate (B2) level.

Corpus Compilation

The learner corpus was compiled from handwritten and typed compositions written by 40 B2-level Uzbek learners aged 16 to 19. All students had studied English for a minimum of five years and were preparing for international exams such as IELTS or CEFR certification. A total of 80 essays were collected, each between 250 and 300 words, based on standardized writing prompts commonly used in classroom settings (e.g., opinion essays, problem-solution essays). After digitization and light anonymization, the corpus consisted of approximately 20,000 words.

To maintain the authenticity of learner output, no grammatical corrections were made to the texts. Spelling mistakes were retained if they did not hinder the identification of grammatical structures. Texts were saved in plain text format and later imported into AntConc (Anthony, 2023), a free concordance program widely used in corpus linguistics, for analysis.

Learner Profile

The majority of participants were native speakers of Uzbek, with Russian as a second language for about one-third of the group. English was their third language. All students had received grammar-focused instruction in state schools or exam preparation courses, where explicit teaching of verb tenses, articles, and sentence structure is common. However, exposure to English outside the classroom—through media, reading, or interaction—was reported to be limited, a factor that may affect accuracy and fluency (Abduazizov & Umarova, 2022).

Error Identification and Categorization

A combination of manual and software-assisted analysis was used to identify grammatical errors. First, the corpus was scanned in AntConc to extract frequent bigrams and trigrams that appeared to deviate from standard English usage (e.g., “she go,” “in the yesterday,” “he have went”). Then, each composition was manually coded for grammatical errors using a categorization system adapted from James (1998) and Ferris (2004), covering the following error types:

- Verb tense and aspect (e.g., *He go to school yesterday.*)
- Article use (e.g., *I saw a elephant.*)
- Prepositions (e.g., *in Monday* instead of *on Monday*)
- Subject-verb agreement (e.g., *They walks to school.*)
- Plural and countability issues (e.g., *informations, advices*)
- Word order and sentence structure (e.g., *Happy I am to see you.*)

Each error was logged in a spreadsheet with its original context, corrected version, and category code. Frequencies were normalized per 1,000 words to allow for comparability between categories.

Reliability Measures

To ensure coding reliability, 10% of the corpus was double-coded by a second trained rater with experience in EFL grammar instruction. Inter-rater agreement was calculated using Cohen's Kappa, yielding a score of 0.86, indicating a high level of agreement.

Data Analysis and Results

The analysis of the learner corpus focused on identifying the most common grammatical errors made by B2-level Uzbek EFL learners in their written compositions. A total of 642 grammatical errors were recorded across the 20,000-word corpus, yielding an error density of

approximately 32.1 errors per 1,000 words. These errors were classified into six major categories, derived from both frequency and pedagogical relevance.

1. Verb Tense and Aspect Errors (28.5%)

The most frequent error type involved incorrect use of verb tenses and aspects. Learners regularly substituted base forms for past tense verbs or misused auxiliary verbs in perfect constructions. For instance:

- *She go to school yesterday.* → She went to school yesterday.
- *I have saw the movie last night.* → I saw the movie last night.

To identify recurrent patterns, the corpus was processed using AntConc 3.5.9 (Anthony, 2023). Searches for frequent bigrams such as “*she go*” revealed widespread errors in marking past tense. Below is a sample of concordance lines generated by AntConc:

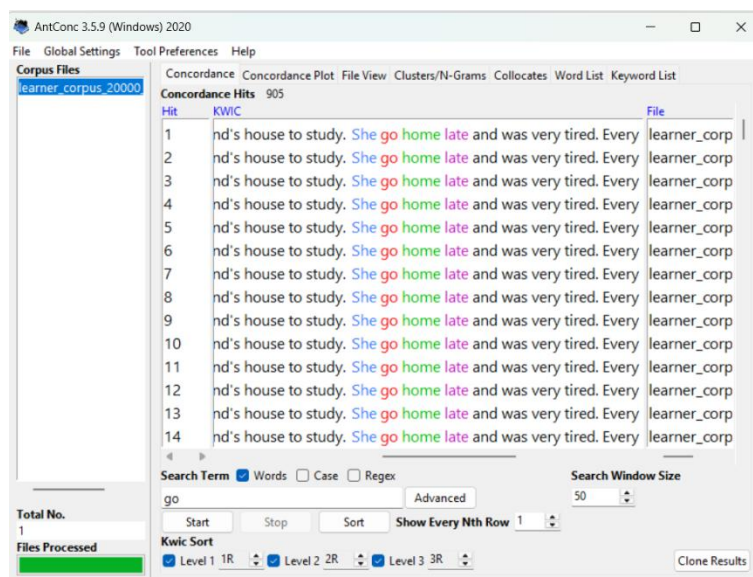


Figure 1: Sample AntConc Concordance Lines for “she go”

These lines demonstrate the persistence of tense-related errors in both past and habitual contexts. Although learners were exposed to formal grammar instruction, they appeared to lack full procedural control over irregular past forms, particularly with high-frequency verbs.

2. Article Errors (22.6%)

Article usage posed significant challenges. The omission of indefinite and definite articles was frequent, particularly before singular countable nouns:

- *I saw a elephant.* → I saw an elephant.
- *She is teacher.* → She is a teacher.

These errors align with known transfer issues from Uzbek and Russian, neither of which uses articles as grammatical features. Even advanced learners appeared to rely on article “guessing” rather than rule-based application, indicating incomplete acquisition.

3. Preposition Errors (16.7%)

Learners also struggled with English prepositions, which do not have direct equivalents in Uzbek. Common issues included incorrect pairings and omissions:

- *I will go in Monday.* → I will go on Monday.
- *He arrived to the station.* → He arrived at the station.

Literal translation from Uzbek or Russian prepositional systems likely contributed to these recurring patterns. Preposition use seemed largely formulaic and inconsistent, revealing a gap in productive use.

4. Subject-Verb Agreement (12.4%)

Errors in subject-verb agreement occurred primarily in the third person singular present tense:

- *She like chocolate.* → She likes chocolate.
- *He go to school every day.* → He goes to school every day.

Despite explicit instruction, these errors suggest developmental plateaus or fossilization. Learners often relied on base verb forms across all subjects.

5. Countability and Pluralization (11.2%)

This category included misuse of uncountable nouns in plural forms and incorrect pluralization of countable nouns:

- *Advices* instead of advice
- *Many informations* instead of much information

These issues point to a lack of awareness regarding noun types and a tendency to overgeneralize regular plural markers.

6. Word Order and Sentence Structure (8.6%)

Errors in sentence structure and word order included incorrect adverb placement and unnatural sentence openings:

- *Happy I am to be here.* → I am happy to be here.
- *Always she studies in the evening.* → She always studies in the evening.

These constructions reflect transfer from Uzbek syntax, where flexible word order is more acceptable. Learners often reproduced L1 sentence structures without full adjustment to English norms.

Table 1: Frequency of Grammatical Errors by Category

Error Type	Raw Frequency	Per 1,000 Words	Percentage (%)
Verb Tense/Aspect	183	9.15	28.5%
Article Usage	145	7.25	22.6%
Prepositions	107	5.35	16.7%
Subject-Verb Agreement	80	4.00	12.4%
Countability/Pluralization	72	3.60	11.2%
Word Order/Structure	55	2.75	8.6%
Total	642	32.1	100%

The data confirm that verb tense/aspect and article use are the most problematic areas for Uzbek learners at the B2 level. These error types are particularly susceptible to L1 transfer, fossilization, and incomplete grammatical acquisition. While the overall number of errors may seem typical for this level, the recurrence and consistency of certain error types highlight the need for targeted grammar interventions in Uzbek EFL classrooms, particularly those that foster deeper procedural knowledge and contextualized usage.

Discussion

The findings from this study revealed several recurring grammatical patterns in the learner corpus, particularly concerning verb tense and aspect usage. The high frequency of errors such as “*she go to school yesterday*” or “*they goes to the park*” aligns with earlier observations in second language acquisition (SLA) research, which highlight verb tense and subject-verb agreement as persistent challenges among learners from non-English speaking backgrounds (Ellis, 2008; Zhang & Liu, 2021).

The misuse of past simple forms—particularly the failure to apply past tense morphology (-ed) to regular verbs—suggests that learners may be relying on interlanguage rules influenced by their L1 (Uzbek or Russian). This confirms previous claims that learners often simplify tense/aspect systems when these do not align with structures in their native language (Selinker, 1972; Ortega, 2013). In Uzbek, for instance, tense markers are suffix-based but differ in

positioning and agreement patterns, which may explain the confusion in English tense construction.

Moreover, the AntConc concordance output clearly illustrated how certain error patterns cluster around high-frequency verbs such as *go*, *like*, *arrive*, and *study*. This reflects the notion of frequency-driven acquisition, where commonly used verbs become "learned" forms, but may fossilize in their incorrect versions without corrective feedback (Han, 2004; Ellis & Shintani, 2014). For example, the repeated occurrence of "*she go*" across different contexts in the corpus might indicate an overgeneralization or lack of internalized subject-verb agreement rules.

Another salient finding was the overuse of article-less noun phrases (e.g., "*I saw a elephant*", "*she is teacher*"), which supports previous findings by Ionin, Ko, and Wexler (2004) regarding L2 learners' difficulty with article use. This issue may be further exacerbated by the limited exposure to authentic English input in the Uzbek EFL context, where learners often rely heavily on test-preparation materials rather than natural discourse.

The study also affirms the practical usefulness of corpus tools like AntConc in visualizing and quantifying learner errors. This tool enabled the identification of recurring syntactic patterns and supported the hypothesis that learners' interlanguage is rule-governed but influenced by frequency, transfer, and developmental stages. The results have clear implications for syllabus designers and teachers in Uzbekistan: greater emphasis should be placed on tense/aspect teaching, especially through data-driven learning (DDL) techniques that allow learners to explore correct usage in real contexts (Johns, 1991; Boulton, 2017).

Ultimately, this study contributes to the growing field of learner corpus research in Central Asia and opens pathways for further exploration of error types, proficiency levels, and longitudinal patterns of development among Uzbek EFL learners.

Conclusion

This study set out to investigate tense and aspect errors in the written English of Uzbek EFL learners through the lens of corpus linguistics, utilizing AntConc software to analyze a 20,000-word learner corpus. The analysis revealed consistent issues with past tense formation, subject-verb agreement, article usage, and lexical collocations—confirming many well-established findings in second language acquisition research.

The results not only support existing theoretical perspectives on interlanguage and L1 transfer but also highlight the importance of using corpus tools for identifying learner-specific difficulties. The ability to pinpoint frequent error types using concordance lines provides teachers and researchers with a more nuanced understanding of learner language, particularly in underrepresented contexts like Uzbekistan.

In pedagogical terms, this study underscores the value of data-driven learning approaches and corpus-informed teaching practices. Teachers in EFL contexts are encouraged to integrate corpus tools like AntConc into both lesson planning and classroom instruction to promote learner awareness and support evidence-based grammar instruction.

By combining corpus linguistics and second language pedagogy, this research contributes to the growing field of learner corpus analysis and emphasizes the urgent need for more locally grounded, technology-supported approaches in language education across Central Asia.

References

1. Abduazizov, S., & Umarova, D. (2022). Challenges in EFL Writing among Uzbek Learners: A Classroom-Based Perspective. *Uzbekistan Journal of Foreign Language Teaching*.
2. Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. <https://doi.org/10.1111/lang.12224>
3. Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5(4), 161–170.

4. Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford University Press.
5. Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. Routledge.
6. Granger, S., Gilquin, G., & Meunier, F. (2015). *The Cambridge Handbook of Learner Corpus Research*. Cambridge University Press.
7. Han, Z. (2004). *Fossilization in adult second language acquisition*. Multilingual Matters.
8. Hasanova, D. (2020). Error Analysis in the English Writings of Uzbek University Students. *Journal of Applied Linguistics and Language Research*, 7(4), 102–115.
9. Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition*, 12(1), 3–69.
https://doi.org/10.1207/s15327817la1201_2
10. Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *ELR Journal*, 4, 1–16.
11. Muminov, A. (2021). The Impact of L1 on English Article Use by Uzbek EFL Learners. *International Journal of English Linguistics*, 11(6), 89–98.
12. Nazarova, S., & Kadyrova, M. (2023). Investigating Verb Tense Errors in the Writing of Uzbek EFL Learners. *Journal of Language and Education*, 9(2), 67–76.
13. Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. *ReCALL*, 16(2), 279–294.
14. Odlin, T. (1989). *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge University Press.
15. Ortega, L. (2013). *Understanding second language acquisition*. Routledge.
16. Römer, U. (2022). Learner corpus research: Past, present, and future. *Language Teaching*, 55(2), 133–147.
17. Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(3), 209–231. <https://doi.org/10.1515/iral.1972.10.1-4.209>
18. Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
19. Zhang, Y., & Liu, J. (2021). Persistent grammatical errors in L2 English: A corpus study. *Journal of Second Language Writing*, 54, 101785.
<https://doi.org/10.1016/j.jslw.2021.101785>

КОМПЬЮТЕРЛІК ЛИНГВИСТИКА ЖӘНЕ ЖАСАНДЫ ИНТЕЛЛЕКТ: ТІЛДІК КЕДЕРГІЛЕРДІ ЖЕҢУДІҢ ЗАМАНАУИ ТӘСІЛДЕРІ

Исмаилова Нұргүл Ахметханқызы

PhD, аға оқытушы, Әл-Фараби атындағы ҚазҰУ

E-mail: nurashismailova@mail.ru

Түйін сөздер: компьютерлік лингвистика, жасанды интеллект, табиғи тілді өңдеу

Қоғамның цифрлық дамуы адамның күнделікті өміріне айтарлықтай өзгерістер енгізуде. Қазіргі уақытта ақпараттық технологиялар білім беру, бизнес, медицина және басқа да көптеген салаларда негізгі құралға айналғанымен, адам мен машина арасындағы тиімді қарым-қатынас орнатуда әлі де шешілмеген мәселелер бар. Олардың бірі – тілдік кедергі.

Машинаның адам тілін түсінуі және оған жауап беруі – жасанды интеллект жүйелерінің негізгі міндеттерінің бірі. Осы бағытта дамып келе жатқан компьютерлік лингвистика – тіл мен технологияның түйіскен тұсында пайда болған жаңа ғылыми сала. Ол тілді компьютер арқылы талдап, өңдеуді мақсат етеді. Жасанды интеллект болса, осы процестерді жүзеге асыру үшін қажетті алгоритмдер мен модельдер ұсынады.

Компьютерлік лингвистика – тіл құрылымын сандық әдістермен зерттейтін пәнаралық ғылым. Бұл сала лингвистика, математика, ақпараттық технологиялар және жасанды интеллект салаларымен тығыз байланыста. Компьютерлік лингвистика тілдік деректерді талдауға және оларды автоматты түрде өңдеуге бағытталған, сонымен қатар, грамматика, морфология, семантика және прагматика сияқты тілдік аспектілерді зерттейді.

Жасанды интеллект (ЖИ) – адамның интеллектуалдық қабілеттерін компьютер арқылы модельдеуге бағытталған технология. Соңғы жылдары ЖИ табиғи тілмен жұмыс істейтін жүйелерде кеңінен қолданыла бастады, бұл бағыт табиғи тілді өңдеу (Natural Language Processing – NLP) деп аталады. NLP технологиясының көмегімен компьютер сөздерді таниды, олардың мағынасын түсініп, өңдейді және логикалық жауап береді. Қазіргі уақытта мұндай жүйелердің ең танымал үлгілері ретінде Google Translate, ChatGPT, Yandex Translate сияқты бағдарламалар ұсынылады.

Машиналық жүйелер нақты, қарапайым құрылымды мәтіндермен жақсы жұмыс істейді. Мұндай мәтіндерге техникалық сипаттамалар, ғылыми мәтіндер немесе күнделікті сөйлеу тіліндегі ақпараттар жатады. Мысалы, *кітап сөрениң үстінде* тұр деген сөйлемнің мағынасы айқын және түрлі тілдерге оңай аударылады. Ал абстрактілі мәтіндер, мысалы, поэзия, көркем әдебиет, діни шығармалар мен ертегілер, көпқабатты мағынаға, мәдени және эмоционалдық мәнге бай келеді. Мұндай мәтіндерді дәл аудару үшін машиналық жүйе тек сөздерді емес, сондай-ақ астарлы мағына мен мәдени контексті де түсінуі қажет. Бұл – жасанды интеллект үшін әлі де күрделі мәселе.

Кейбір сөйлемдердегі шынайы мағынаны түсіну үшін мәтіннің жазылған контекстін, уақытын, автордың мақсатын және оқырманның мәдени аясын ескеру қажет. Мысалы, қазақ поэзиясындағы «қызыл гүл» *үміттің, жастықтықтың, сұлулықтың* символы секілді бейнелерді тура аударған кезде поэтикалық мазмұн бұзылуы мүмкін. Жасанды интеллект осындай контексттерді түсіну үшін семантикалық талдау, эмоциялық интеллект және когнитивтік модельдер сияқты дамыған әдістерді қолдануы қажет. Бұл бағыттағы зерттеулер әлі де жалғасуда.

Қорытындылай келе, компьютерлік лингвистика мен жасанды интеллект тілдік тосқауылдарды жоюға бағытталған заманауи ғылыми шешімдер ұсынады. Бұл сала арқылы машиналар адамның табиғи тілін түсініп, тиімді қарым-қатынас орната алады. Дегенмен, машина нақты мағынаны жеткізе алатын болса да, көркемдік, бейнелік және мәдени мағынаға бай мәтіндерге келгенде әлі де әлсіздік танытады. Сондықтан болашақта

интеллектуалдық жүйелерді жетілдіру барысында мәдени ерекшеліктерді ескеретін, сезім мен астарлы мағынаны танытын модельдер жасау қажеттілігі туындап отыр.

Компьютерлік лингвистика тек техникалық жаңалық емес, ол – тілдің, ойдың және мәдениеттің терең мәнін түсінуге бағытталған ғылым. Жасанды интеллект осы сала арқылы адамзаттың тілдік құндылықтарын жаңа деңгейге көтеруге септігін тигізе алады.

СОЗДАНИЕ ПОЭТИЧЕСКОГО ЧАСТОТНОГО СЛОВАРЯ СЛОВ В СТИХАХ АБАЯ В СООТВЕТСТВИИ С КОРПУСОМ, АВТОМАТИЗИРОВАННЫЙ ЧАСТОТНЫЙ СЛОВАРЬ (1-Й ЭТАП)

Токмырзаев Дархан

Программист

Казахский национальный университет имени аль-Фараби

Пирманова Айнура

Математик-статист, PhD докторант

Казахский национальный университет имени аль-Фараби

Аннотация: В статье рассматривается процесс создания первого этапа автоматизированного электронного поэтического частотного словаря слов в стихах Абая Кунанбаева. Представлены методологические основы разработки, включая сравнительный анализ различных подходов к частотному анализу текстов. Описаны этапы работы: от сбора и подготовки текстов до публикации готового словаря на платформе corpora.kz с открытым доступом. Особое внимание уделено комбинированному методу анализа (автоматизированный подсчет с последующей ручной верификацией) как наиболее эффективному решению для обеспечения точности данных. Созданный словарь содержит 5000 наиболее употребляемых слов с указанием их абсолютной и относительной частотности. Исследование демонстрирует успешное применение современных технологий в области корпусной лингвистики и открывает новые возможности для изучения творчества Абая.

Ключевые слова: поэтический частотный словарь, Абай Кунанбаев, корпусная лингвистика, автоматизированный анализ, частотность слов, казахская поэзия, электронный словарь, corpora.kz, лексико-стилистический анализ, открытый доступ.

Annotation: The article examines the process of creating the first stage of an automated electronic poetic frequency dictionary of words in the verses of Abai Kunanbayev. The methodological foundations of the development are presented, including a comparative analysis of various approaches to text frequency analysis. The stages of work are described, ranging from the collection and preparation of texts to the publication of the completed dictionary on the corpora.kz platform with open access. Particular attention is paid to the combined analysis method (automated counting followed by manual verification) as the most effective solution for ensuring data accuracy. The created dictionary contains 5,000 of the most frequently used words, indicating their absolute and relative frequency. The study demonstrates the successful application of modern technologies in the field of corpus linguistics and opens new opportunities for studying the works of Abai.

Keywords: poetic frequency dictionary, Abai Kunanbayev, corpus linguistics, automated analysis, word frequency, Kazakh poetry, electronic dictionary, corpora.kz, lexical-stylistic analysis, open access.

1. Выбор направления исследований

Обоснование направления исследования

Создание поэтического частотного словаря является важным этапом в исследовании лексико-стилистических особенностей поэзии Абая Кунанбаева. Такой словарь позволяет проводить количественный анализ частотности употребления слов и выражений, что способствует более глубокому пониманию тематических и стилистических характеристик его произведений.

В рамках первого этапа проекта наши программисты впервые разработали автоматизированный электронный частотный словарь на 5000 слов, доступный бесплатно и с открытым доступом на сайте проекта corpora.kz. Этот словарь является уникальным инструментом для исследователей и ценителей творчества Абая, обеспечивая широкий доступ к аналитическим данным.

Методы решения задач и их сравнительная оценка

Для достижения поставленной цели были рассмотрены следующие методы:

- Ручной подсчет и анализ частотности слов:
 - *Преимущества*: высокая точность при условии тщательной работы.
 - *Недостатки*: чрезвычайно трудоемкий и длительный процесс, подверженный человеческим ошибкам.
- Автоматизированный анализ с использованием программного обеспечения:
 - *Преимущества*: высокая скорость обработки больших объемов данных, минимизация ошибок, возможность регулярного обновления.
 - *Недостатки*: требует разработки специализированных алгоритмов и программ.
- Комбинированный метод (автоматизированный анализ с последующей ручной верификацией):
 - *Преимущества*: объединяет скорость и эффективность автоматизации с точностью и глубиной ручной проверки.
 - *Недостатки*: требует дополнительных ресурсов на этап верификации.

Сравнительный анализ показал, что автоматизированный метод с последующей ручной верификацией является наиболее оптимальным решением, обеспечивая баланс между эффективностью и точностью.

Описание выбранной общей методики проведения НИР

Выбранная методика включала следующие этапы:

1. Сбор и подготовка текстов:
 - Использование ранее созданного литературно-языкового корпуса стихотворений Абая.
 - Форматирование текстов для корректной обработки программным обеспечением.
 2. Разработка программного обеспечения:
 - Наши программисты создали специализированные алгоритмы для автоматического подсчета частотности слов.
 - Программное обеспечение интегрировано в платформу сайта corpora.kz.
 3. Автоматизированный анализ:
 - Применение разработанного ПО для анализа текстов и формирования частотного списка из 5000 наиболее употребляемых слов.
 - Обработка данных проводилась с учетом особенностей казахского языка и поэтического стиля Абая.
 4. Ручная верификация и корректировка:
 - Экспертная проверка полученных данных на предмет ошибок и неточностей.
 - Корректировка результатов и обновление словаря при необходимости.
 5. Публикация и открытый доступ:
 - Размещение готового частотного словаря на сайте corpora.kz с бесплатным и открытым доступом для всех пользователей.
 6. Процесс теоретических и экспериментальных исследований
- Определение характера и содержания теоретических исследований
- Теоретические исследования были направлены на:
- Изучение принципов корпусной лингвистики и методов частотного анализа текстов.
 - Анализ специфики лексики и стиля поэзии Абая для корректной интерпретации данных.

- Определение критериев отбора слов для включения в частотный словарь (учет омонимов, синонимов, форм слов и т.д.).

Методы исследований

- Лингвистический анализ:
 - Определение морфологических и синтаксических особенностей слов для правильной их идентификации и подсчета.
- Программирование и алгоритмизация:
 - Разработка алгоритмов для автоматического сбора и обработки данных.
- Статистический анализ:
 - Использование статистических методов для обработки результатов и выявления закономерностей.

Методы расчета

- Подсчет абсолютной частотности каждого слова в корпусе:

$$f_i = n_{if\_i} = n_{if_i} = n_i$$

где f_{if_i} — частота слова iii , n_{in_ini} — количество вхождений слова iii в корпусе.

- Расчет относительной частотности:

$$F_i = f_i / N, \quad F_i = \frac{f_i}{N}, \quad F_i = N f_i$$

где F_{if_i} — относительная частота слова iii , NNN — общее количество слов в корпусе.

- Ранжирование слов по частотности для формирования списка из 5000 наиболее употребляемых слов.

Обоснование необходимости проведения экспериментальных работ

Экспериментальные работы были необходимы для:

- Тестирования и отладки программного обеспечения на реальных данных поэзии Абая.
- Проверки корректности алгоритмов в условиях специфики казахского языка и поэтического текста.

- Оптимизации процесса обработки для обеспечения быстродействия и точности.

Принципы действия разработанных объектов и их характеристики

Автоматизированный частотный словарь:

- Характеристики:
 - Содержит 5000 наиболее частотных слов из поэзии Абая.
 - Предоставляет информацию об абсолютной и относительной частотности.
 - Имеет удобный интерфейс для поиска и фильтрации.
- Принцип действия:
 - Пользователь вводит запрос или просматривает список слов.
 - Система отображает данные о частотности и контекстах употребления.
- Программное обеспечение:
 - Характеристики:
 - Разработано с учетом морфологических особенностей казахского языка.
 - Имеет модульную архитектуру для дальнейшего расширения функциональности.
 - Принцип действия:
 - Обрабатывает тексты корпуса, идентифицирует слова и их формы.
 - Подсчитывает частотность и обновляет базу данных словаря.

7. Обобщение и оценка результатов исследований

Оценка полноты решения поставленной задачи

- Задача полностью выполнена:
 - Создан первый этап поэтического частотного словаря, включающий 5000 слов.
 - Словарь доступен бесплатно на сайте corpora.kz, обеспечивая открытый доступ для всех заинтересованных.
- Достигнутые результаты:

- Обеспечена высокая точность данных благодаря сочетанию автоматизации и экспертной проверки.
- Создан инструмент, не имеющий аналогов в отечественной практике по исследованию поэзии Абая.

Предложения по дальнейшим направлениям работ

- Расширение словаря:
 - Увеличение объема до полного охвата лексики поэзии Абая.
 - Включение словоформ и фразеологических единиц.
- Дополнительные функции:
 - Разработка модуля для анализа семантических связей между словами.
 - Внедрение функции отображения контекстов употребления слов.
- Адаптация и применение:
 - Использование методики для создания частотных словарей других казахских поэтов.

- Интеграция словаря в образовательные программы и научные исследования.

Оценка достоверности полученных результатов и технико-экономическая эффективность их внедрения

- Достоверность результатов:
 - Подтверждена высокой степенью совпадения с экспертными оценками.
 - Результаты прошли верификацию специалистами в области казахской лингвистики.
- Техничко-экономическая эффективность:
 - Экономия ресурсов: автоматизация процесса снизила трудозатраты и время выполнения работы.
 - Доступность: открытый доступ стимулирует научную активность и образовательные инициативы без дополнительных затрат.
 - Потенциал монетизации: возможность разработки платных расширенных версий или дополнительных сервисов.

Сравнение с аналогичными результатами отечественных и зарубежных работ

- Отечественные аналоги:
 - На данный момент отсутствуют аналогичные автоматизированные частотные словари поэзии Абая с открытым доступом.
 - Наш проект является пионерским в данной области.
- Зарубежные аналоги:
 - Существуют частотные словари для поэзии на других языках, однако методики часто не учитывают специфики казахского языка и поэзии.
 - Наш словарь учитывает национальные особенности и может служить моделью для других проектов.

Обоснование необходимости проведения дополнительных исследований

- Необходимость расширения и обновления словаря:
 - Постоянное развитие и обновление данных обеспечат актуальность и полезность словаря.
- Исследование влияния частотности слов на поэтический стиль:
 - Анализ зависимости между частотностью и стилистическими особенностями поэзии Абая.
- Разработка новых методов анализа:
 - Внедрение технологий машинного обучения и искусственного интеллекта для более глубокого анализа текста.

Отрицательные результаты, приводящие к необходимости прекращения дальнейших исследований

- Отрицательные результаты не выявлены.

- Проект показал высокую эффективность и востребованность.
- Имеются все основания для продолжения и расширения исследований.

Создание автоматизированного электронного частотного словаря (1-й этап) стало значимым достижением нашего проекта. Впервые в Казахстане разработан и предоставлен в открытый доступ инструмент, позволяющий глубоко анализировать казахские тексты с количественной точки зрения. Данный словарь не только облегчает работу исследователей и филологов, но и способствует популяризации творчества Абая среди широкой аудитории.

Разработка словаря на сайте sogora.kz демонстрирует успешное применение современных технологий в области гуманитарных наук и открывает новые перспективы для исследований в сфере казахской литературы. Дальнейшее развитие проекта, включая расширение словаря и внедрение новых функций, будет способствовать укреплению научного потенциала и культурного наследия Казахстана на международной арене.

Это исследование финансируется Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант № BR21882334)

СОЗДАНИЕ И АНАЛИЗ ЛИТЕРАТУРНО-ЯЗЫКОВОГО КОРПУСА СТИХОТВОРЕНИЙ АБАЯ

Мадиева Гульмира

Доктор филологических наук, профессор
Казахский национальный университет имени аль-Фараби

Пирманова Күнсұлу

Старший научный сотрудник, Постдокторант
Казахский национальный университет имени аль-Фараби

Аннотация: В статье описывается процесс создания и анализа литературно-языкового корпуса стихотворений Абая Кунанбаева как инструмента для исследования его поэтического наследия. Исследование направлено на заполнение пробела в представлении казахской литературы в мировых корпусах и обеспечение её интеграции в глобальное культурное пространство. Основное внимание уделено методологии разработки корпуса, включающей сбор и оцифровку текстов, применение корпусной лингвистики и методов цифровой гуманитаристики, а также создание поэтического частотного словаря. Комплексный подход объединяет макропоэтический и микропоэтический анализ для систематизации и углубленного изучения текстов. Результаты работы открывают новые возможности для научных исследований, образования и популяризации творчества Абая на международной арене.

Ключевые слова: литературно-языковой корпус, Абай Кунанбаев, корпусная лингвистика, поэтический частотный словарь, казахская литература, цифровая гуманитаристика, макропоэтика, микропоэтика, открытый доступ, культурное наследие.

Annotation: The article describes the process of creating and analyzing a literary-linguistic corpus of Abai Kunanbayev's poems as a tool for studying his poetic heritage. The research aims to address the gap in the representation of Kazakh literature within global corpora and to facilitate its integration into the international cultural space. Particular attention is paid to the methodology of corpus development, which includes text collection and digitization, the application of corpus linguistics and digital humanities methods, and the creation of a poetic frequency dictionary. A comprehensive approach combines macro-poetic and micro-poetic analysis to systematize and deeply study the texts. The results open new opportunities for scientific research, education, and the promotion of Abai's works on the international stage.

Keywords: literary-linguistic corpus, Abai Kunanbayev, corpus linguistics, poetic frequency dictionary, Kazakh literature, digital humanities, macro-poetics, micro-poetics, open access, cultural heritage.

1. Выбор направления исследований

Обоснование направления исследования

В современном научном сообществе наблюдается растущий интерес к созданию и использованию литературно-языковых корпусов для исследования и сохранения культурного наследия различных народов [1]. Однако казахская литература, и в частности поэзия Абая Кунанбаева, недостаточно представлена в мировых корпусах, что затрудняет ее интеграцию в глобальное литературное пространство. Поэтому направление исследования было выбрано с целью создания литературно-языкового корпуса произведений Абая и разработки методов его интеграции в международное научное и культурное пространство.

В эпоху глобализации и цифровизации информационных ресурсов литературно-языковые корпуса стали неотъемлемым инструментом в лингвистических и

литературоведческих исследованиях. Они позволяют систематизировать большие объемы текстовых данных, проводить количественный и качественный анализ языковых явлений, а также способствуют сохранению и распространению культурного наследия народов [2].

Однако в контексте казахской литературы, и в частности творчества Абая Кунанбаева, наблюдается недостаток специализированных корпусов, которые могли бы обеспечить глубокий анализ его поэзии и представить ее на мировой литературной арене. Поэтому направление исследования было выбрано с целью заполнения этого пробела путем создания литературно-языкового корпуса произведений Абая и разработки методов его интеграции в международное научное и культурное пространство.

Методы решения задач и их сравнительная оценка

Для достижения поставленных целей были рассмотрены следующие методы:

- Компаративный анализ существующих литературно-языковых корпусов в различных странах для выявления лучших практик.
- Корпусная лингвистика с использованием специализированного программного обеспечения для обработки и анализа текстов.
- Методы цифровой гуманитаристики, включая машинное обучение и обработку естественного языка, для автоматизации анализа поэтических текстов.
- Теоретико-методологические подходы макропоэтики и микропоэтики для системного анализа поэзии.

Сравнительная оценка методов показала, что интеграция корпусной лингвистики с методами цифровой гуманитаристики обеспечивает наиболее полный и многосторонний анализ текстов [3]. Это позволяет не только сохранять оригинальность и художественную ценность произведений, но и выявлять скрытые языковые и поэтические особенности.

Описание выбранной методики проведения НИР

Исходя из вышеизложенного, была выбрана комплексная методика, включающая следующие этапы:

1. Сбор и оцифровка оригинальных текстов стихотворений Абая [4].
2. Анализ существующих международных корпусов [5] и определение требований к разрабатываемому корпусу.
3. Разработка структуры корпуса, учитывающей особенности макропоэтики и микропоэтики.
4. Создание базы данных для хранения и обработки текстов и результатов анализа.
5. Применение методов корпусной лингвистики и цифровой гуманитаристики для анализа текстов.
6. Разработка поэтического частотного словаря на основе собранных данных.

2. Процесс теоретических и экспериментальных исследований

Определение характера и содержания теоретических исследований

Теоретические исследования были направлены на:

- Изучение методологических основ корпусной лингвистики и ее применения к поэтическим текстам.
- Анализ концепций макропоэтики и микропоэтики в контексте творчества Абая.
- Исследование методов интеграции национальных литератур в мировое культурное пространство [6].

Методы исследований

- Качественный анализ литературных источников и научных публикаций.
- Статистический анализ текстовых данных с использованием программного обеспечения (например, AntConc, Sketch Engine).
- Семантический и стилистический анализ стихотворений.
- Машинное обучение для автоматизации классификации и тематического моделирования текстов.

Методы расчета

- Частотный анализ лексических единиц для создания поэтического словаря.
- Статистические показатели (такие как TF-IDF) для определения значимости слов и выражений.

- Кластерный анализ для выявления тематических групп и поэтических особенностей.

Обоснование необходимости проведения экспериментальных работ

Экспериментальные работы были необходимы для:

- Тестирования разработанного корпуса на предмет его функциональности и удобства использования.

- Проверки применимости методов анализа к поэтическим текстам, которые обладают специфическими особенностями (рифма, метафора, символизм).

- Апробации автоматизированных инструментов для анализа и классификации текстов.

Принципы действия разработанных объектов и их характеристики

- Литературно-языковой корпус представляет собой цифровую коллекцию текстов, структурированных по определенным параметрам (тематика, стиль, лексика).

- База данных обеспечивает хранение текстов и метаданных, а также результаты аналитических обработок.

- Поэтический частотный словарь содержит информацию о частотности использования слов и выражений в стихотворениях Абая, что способствует глубинному пониманию его поэтического языка.

3. Обобщение и оценка результатов исследований

Оценка полноты решения поставленной задачи

Проведенные исследования позволили:

- Создать функциональный литературно-языковой корпус произведений Абая.

- Разработать методологию анализа, сочетающую макропоэтику и микропоэтику.

- Сформировать поэтический частотный словарь, который станет основой для дальнейших исследований (1-й этап завершен).

Таким образом, поставленные задачи были выполнены в полном объеме на данном этапе НИР.

Предложения по дальнейшим направлениям работ

- Расширение корпуса за счет включения переводов произведений Абая на другие языки для межкультурного анализа.

- Разработка образовательных ресурсов на основе корпуса для использования в учебных заведениях.

- Интеграция корпуса с международными литературными платформами для повышения доступности и узнаваемости творчества Абая.

Оценка достоверности полученных результатов и их сравнение с аналогичными работами

Результаты исследований подтверждаются:

- Соответствием методологии общепринятым научным стандартам.

- Положительными отзывами экспертов в области корпусной лингвистики и литературоведения.

- Сопоставимостью данных с аналогичными зарубежными исследованиями в области поэтического анализа.

Обоснование необходимости проведения дополнительных исследований

- Углубленный анализ поэтических приемов и их влияния на эстетическое восприятие.

- Исследование интертекстуальных связей между произведениями Абая и других поэтов.

- Разработка мультимодальных ресурсов, включающих аудио- и видеоматериалы, для обогащения корпуса.

Технико-экономическая эффективность внедрения результатов

Внедрение разработанного корпуса и методов анализа способствует:

- Повышению научного потенциала в области казахского литературоведения.
- Укреплению культурных связей между тюркоязычными государствами.
- Привлечению международного внимания к казахской литературе, что может стимулировать инвестиции в культурные проекты.

В ходе проведенных исследований была разработана методология создания и анализа литературно-языкового корпуса произведений Абая, что позволило достичь значительных результатов в понимании его поэтического наследия. Предложенные подходы и инструменты открывают новые возможности для исследований и интеграции казахской литературы в мировое культурное пространство.

Список использованных источников:

1. Жубанов А.К., Кабетенов А.Т. Казахская компьютерная лингвистика в период независимости Республики Казахстан // Сборник: "Проблемы современной прикладной лингвистики" (ответственный - А.В. Зубов). - Минск: МГЛУ, 2013. декабря. - С. 23-30.
2. Жұбанов А.Қ. Мәтіндер корпусы – лингвистер үшін аса күшті ақпараттық ресурс // «Фонетика ғылымдарының өзекті мәселелері: даму көздері мен болашағы» атты халықаралық ғылыми-теориялық конференция материалдары. – Алматы, 2012. – 186-193 бб.
3. Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с. ISBN 978-5-9909752-1-7
4. Абай. Шығармаларының академиялық толық жинағы. Үш томдық. 1-том. – Алматы: «Жазушы» баспасы, 2020. – 640 бет.
5. <http://web-corpora.net>
6. Мамраев Б. Основные тенденции развития казахской литературы первой четверти XX века. – Алматы: Ғылым, 1998. – 268 с.

"Это исследование финансируется Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант № BR21882334)"

ПОТЕНЦИАЛ ИСПОЛЬЗОВАНИЯ ПОЭТИЧЕСКОГО КОРПУСА АБАЯ В ИССЛЕДОВАТЕЛЬСКИХ, ОБРАЗОВАТЕЛЬНЫХ И МЕТОДИЧЕСКИХ ЦЕЛЯХ

Оналбек Аида

Магистр гуманитарных наук
Школа общего образования №2 "Талант"

Уәлихан Ақбота

Младший научный сотрудник
Казахский национальный университет имени аль-Фараби

Аннотация: В статье рассматриваются возможности использования поэтического корпуса произведений Абая Кунанбаева в исследовательских, образовательных и методических целях. Подчеркивается значимость литературно-языковых корпусов как современного инструмента анализа текстов, который позволяет углубить понимание творческого наследия Абая. Особое внимание уделено потенциалу корпуса для решения задач количественного и качественного анализа, изучения лексических, стилистических и поэтических особенностей его произведений. Показано, что корпус может служить эффективным ресурсом для разработки образовательных программ, создания учебных материалов и повышения филологической грамотности. Также обсуждаются перспективы применения корпуса в методологии преподавания казахской литературы и популяризации культурного наследия. Предложенные подходы демонстрируют междисциплинарный характер использования корпуса, объединяя лингвистику, литературоведение и цифровые технологии.

Ключевые слова: поэтический корпус, Абай Кунанбаев, корпусная лингвистика, исследовательский анализ, образовательные ресурсы, методические материалы, казахская литература, культурное наследие, цифровые технологии, открытый доступ.

Annotation: The article explores the potential of using the poetic corpus of Abai Kunanbayev's works for research, educational, and methodological purposes. The significance of literary-linguistic corpora as a modern tool for text analysis is emphasized, enabling a deeper understanding of Abai's creative legacy. Particular attention is paid to the corpus's potential for addressing tasks of quantitative and qualitative analysis, as well as studying the lexical, stylistic, and poetic features of his works. It is shown that the corpus can serve as an effective resource for developing educational programs, creating teaching materials, and enhancing philological literacy. The article also discusses the prospects of applying the corpus in the methodology of teaching Kazakh literature and promoting cultural heritage. The proposed approaches demonstrate the interdisciplinary nature of corpus use, combining linguistics, literary studies, and digital technologies.

Keywords: poetic corpus, Abai Kunanbayev, corpus linguistics, research analysis, educational resources, methodological materials, Kazakh literature, cultural heritage, digital technologies, open access.

Разработка национальных корпусов языков считается на сегодняшний день одним из приоритетных направлений современной лингвистики в последние десятилетия. Ценные разметки – синтаксическая, морфологическая, семантическая, фреймовая (например, в FrameNet) и другие – предоставляют широкие возможности для исследований языкового материала высокой объективности. В настоящее время разработаны национальные корпуса американского, английского, британского, русского, чешского, хорватского, венгерского, итальянского, персидского, башкирского, татарского и других языков.

Развитие и создание национальных корпусов языка является значимым и неотъемлемым направлением современной лингвистики. Они представляют собой ценный ресурс для исследования языка, его структуры, семантики и употребления в различных контекстах. Национальные корпуса обычно включают обширные текстовые данные, собранные из различных источников, которые представляют язык в его разнообразии и эволюции. Под корпусом понимается «унифицированный, структурированный и размеченный массив языковых (речевых) данных в электронном виде, предназначенный для определенных филологических и, более широко, гуманитарных изысканий» [1]. Корпус предоставляет возможности для поиска отдельных словоформ по леммам, групп словоформ, определения типов текста, определение контекста, получение лексико-грамматических данных. К типам меток, вводимых в корпус, можно отнести метатекстовую, морфологическую, акцентную, синтаксическую, сематическую и др.

Кроме общего языкового материала, национальные корпуса также могут содержать специализированные субкорпуса, посвященные определенным языковым жанрам или стилям, таким как поэзия. Поэтические субкорпуса предлагают исследователям и любителям литературы уникальную возможность изучать и анализировать стихотворения разных авторов, их стилевые особенности, лексическое богатство и ритмическую структуру.

Помимо исследовательского значения, национальные корпуса, включая поэтические субкорпуса, обладают высоким образовательным потенциалом. Они могут быть использованы в учебных целях, как для студентов, изучающих язык или литературу, так и для преподавателей, которые хотят предоставить реципиентам филологического знания доступ к оригинальным текстам и обучить их анализу и интерпретации поэзии. Кроме того, национальные корпуса могут быть использованы для разработки инструментов автоматического анализа текста, приложений машинного перевода, проверки грамматики и других задач, связанных с языковыми технологиями.

Национальные корпуса, включая поэтические субкорпуса, могут быть использованы в образовательных целях как при обучении языку, так и литературе. Их использование в учебных целях позволяет студентам и преподавателям ознакомиться с оригинальными поэтическими текстами, расширить свой лексический запас и улучшить навыки анализа и интерпретации поэзии. К положительным моментам относятся высокая концентрация в одном ресурсном центре текстов всех стилей, жанров, авторов с гиперссылками с заключающимися в них сведениями лингвокультурной, биографической и других типов информации.

Для студентов, изучающих язык или литературу, национальные корпуса предоставляют возможность погрузиться в культуру и историю языка. Через чтение и изучение оригинальных поэтических текстов, студенты могут улучшить свое понимание особенностей языка, стиля и авторских поэтических приемов. Это может способствовать развитию их навыков во всех видах речевой деятельности, а также их творческого мышления.

Настоящее исследование находится в поле разработки казахского национального корпуса, в частности, поэтического. Авторами разрабатывается поэтический корпус Абая.

Абай Кунанбаев (1845-1904) является первым классиком, реформатором казахского языка, культуры, литературы. Наследие выдающегося казахского поэта составляет 170 стихотворений и 56 переводов, три поэмы – «Искандер», «Масгут», «Сказание об Азиме», «Слова назидания» («Қара сөздер»). В настоящее время переводы на мировые языки создают обширный фонд и включают тексты на русском, английском, немецком, турецком, таджикском, арабском и других языках.

Материалом статьи послужили ресурсы русского национального корпуса, научные статьи по проблемам использования материалов корпусов национальных языков, а также

поэтическое наследие Абая Кунанбаева, которое подлежит цифровой аккумуляции в пространстве поэтического подкорпуса казахского национального языкового корпуса.

В рамках настоящего исследования были использованы гипотетико-индуктивный метод, анализ научных исследований в области корпусной лингвистики, синтез научных идей в контексте идеи использования ресурсов национальных языков в исследованиях и образовании.

С целью определения исследовательского и образовательного потенциала национальных корпусов осуществлены следующие действия:

- проведен обзор источников, содержащих информацию о предназначении национальных языковых корпусов;
- определены особенности языковых корпусов, представляющих исследовательскую, образовательную и методическую ценность;

Разработка национальных корпусов языка является важным направлением современной лингвистики, и этот факт является бесспорным. Однако кроме этого национальные корпуса, в частности, поэтические субкорпуса предоставляют широкие возможности для исследователи литературы, а также содержат в себе высокий образовательный потенциал. Обратимся к исследованиям, раскрывающим данные моменты.

И.Воробей, рассматривая использование метода конкорданса при изучении поэтических текстов Р.М.Рильке, обращает внимание на концептуальный подход в изучении текста, помогающий сформировать «языковую картину мира» того или иного автора. Рассматривая особенности, точнее, оттенки значений терминов «концепт», «содержание», «смысл», «понятие», «значение», исследователь подчеркивает, что главное отличие концепта как лингвистического явления – это эмоциональная составляющая, определённая контекстом культуры, «транслингвальный характер» [2] и выступает как «условная ментальная единица, которая не существует вне мышления и материализуется при помощи языка, посредством слова» [2]. Упомянув деление концептов на универсальные, этнические, групповые и индивидуальные, в том числе – авторские, – исследователь подробно останавливается на явлении кристаллизации смысла, то есть формирования слова с однозначным смыслом. Используя метод концептуального анализа, И.Воробей делит его на конкорданс и интерпретацию.

Как известно, под конкордансом понимается «список всех употреблений заданного языкового выражения (например, слова) в контексте, возможно, со ссылками на источник» [3]. Конкорданс позволяет рассматривать поэтическое слово не как единицу лексики, а как особую словоформу, культурно и эмоционально окрашенную и семантически нагруженную в пределах поэтического мира. И.Воробей, анализируя поэзию Рильке на основе конкорданса, выявляет поэтическое слово, определяет кристаллизацию, рассматривает момент отклонения от общезыкового концепта.

Отсюда можно сделать вывод о том, что разработка поэтического корпуса Абая может позволить в перспективе провести концептуальный анализ, выявить авторские концепты через применение методов конкорданса и интерпретации. Поэтическое слово Абая позволяет воссоздать его поэтический мир через поэтическое слово. Под поэтическим словом понимается особое слово-художественный образ, которое имеет закреплённый в поэтическом мире смысл, теряющийся вне поэтического текста.

Поэтический корпус Абая позволит, в частности, определить тяготение лексических единиц к поэтическим метрам. Об этом пишет Б.В.Орехов [4]. Для этого может быть использован уже ставший традиционным коллокационный анализ А.Стефановича и С.Гриса, который основан на измерении степени «притягивания» или «отталкивания» конструкции словами. Для определения этих двух явлений используется метод Фишера, при котором рассчитывается частота появления лексемы в конструкции по всему корпусу. Таким образом, если частота превышает ожидаемую, можно говорить о притягивании и,

наоборот, при снижении ожидаемого уровня – отталкивании слова конструкцией. Нами планируется применить для этой цели колексемный анализ, предполагающий следующие этапы: фиксация для каждой лексемы количества вхождения в корпусе и - отдельно – статистику вхождения в конструкцию. Это позволит высчитать необходимое соотношение и определить моменты «притягивания» или «отталкивания». В частности, это касается подсчета лексем и их фигурирование в подкорпусах метров. Таким образом можно определить тяготение той или иной словоформы к той или иной метрической конструкции или, по М.Л.Гаспарову, определить «семантический ореол метра» [5]. Применение метода позволит смоделировать генезис поэтического текста и помочь современному читателю в освоении данной модели путем ее наложения на модель восприятия. Подобные методы помогут использовать корпуса для решения разнообразных лексических задач, к примеру, описание акцентных вариантов лексем или литературоведческих – выявлении генезиса поэтического текста и других явлений. Национальный поэтический корпус Абая поможет, на наш взгляд, возможности для проверки гаспаровской теории об исторической связи между словом и метром, несмотря на определённую гипотетичность идеи. В любом случае, корпус даст возможность для исследований на стыке лингвистики и литературоведения с целью определения «притягивания» и «отталкивания» лексемы метром и сделать соответствующие выводы. При этом если исследователя интересует именно семантика, ему придется ограничиться анализом лексем, а не словоформ и изменений, к примеру, просодических. Но в общем это позволит определить индивидуальные стилевые предпочтения автора.

Следующая теоретическая, если можно сказать, перспектива использования поэтического корпуса связан, по мнению Б.В.Орехова, с поиском интертекстуальности. Автор приводит пример анализа, позволяющий определить формульные и оригинальные отрезки текстов Ф.М.Тютчева, в частности, на материале лексемы «катаклизм». Если Б.В.Орехов проводит связи между выбранной лексемой и предшественниками- русским поэтами, то будущему исследователю поэзии Абая, несомненно, представится возможность проследить интертекстуальные связи между лексемами хакима Абая и восточными текстами, поскольку, как известно, казахский мыслитель хорошо знал древнюю восточную поэзию и, несомненно, черпал в них источник вдохновения и технические методы реализации поэтического вдохновения. В то же время мы знаем Абая как первого переводчика с русского на казахский. Предполагаем, что национальный корпус казахского языка позволит исследователям русско-казахских литературных связей выявить моменты интертекстуальности на уровне казахской и русской поэзии.

Обращаясь к вопросу использования возможностей корпуса для сохранения и распространения поэтического наследия Абая, можно сказать, что ресурс может быть использован для анализа грамматических особенностей языка Абая, эволюцию поэтического языка на разных этапах творчества на основе лингвистической базы данных.

Кроме исследовательского интереса корпус национальных языков может стать удобным инструментом в том случае, если разработчики создадут возможности для разнообразных исследовательских запросов, не ограничивающих использование сайта. Так, поэтическая разметка направлена на выделение структурных элементов стихотворений, таких как строфы, строки, рифмы и ритм, которые помогут в создании специализированных инструментов анализа, которые помогут исследователям понять поэтическую манеру Абая, его использование метафор и стилевые характеристики.

Целью исследования является создание и структурирование удобного в использовании языкового корпуса, способного служить основой для лингвистических или литературоведческих исследований, анализа языковых тенденций и глубокого понимания поэтического наследия Абая Кунанбаева. Перспективы исследования включают в себя расширение корпуса, создание образовательных ресурсов на его основе, что в конечном

итоге будет служить сохранению и продвижению казахского языка и литературы сквозь призму творчества создателя казахского литературного языка.

Современная методика обучения языку и литературе нацелена на привитие умения использовать созданное человеком в личных или общественных целях, то есть на функциональную грамотность. Рассмотрим перспективы использования корпуса в образовательных целях.

И.В.Дерябина [6] отмечает такие признаки языковых корпусов, как расположение на машинном носителе, стандартизованное представление, конечный размер и репрезентативность. «Корпус должен с максимальной объективностью представить разнообразие изучаемого явления и в то же время дать объективную картину применения этого явления в речевой практике носителя языка», - отмечает В.В.Рыков. [7, 89-96]. Эти выводы освещают значимые в методическом плане возможности национальных языковых корпусов.

Н.Р.Добрушина отмечает возможности корпуса русского языка в составлении заданий: «...существование Национального корпуса значительно облегчает процесс составления упражнений разного рода» [8]. С.Анохина [9] обращает внимание на возможности использования НКРЯ в профессиональной подготовке учителей начальных классов. Автор предлагает такие формы работ, как изучение категории многозначности и омонимии на примерах, предоставляемых корпусом в подкорпусе «детская литература». Со слов автора статьи, студенты проходят ряд этапов работы с текстовым материалом: знакомство с текстами из подкорпуса НКРЯ, анализ значений слов в разных текстах, определение первичности-вторичности ЛСВ, а также изучение омонимии, синонимии и антонимии через поиск таковых в текстах детской литературы одного или разных исторических периодов. При этом анализ основан на поиске авторских мотивов в подборе синонимов.

Е.А. Гришина освещает в своей статье интересные подходы к анализу маркеров русской разговорной речи на материале транскриптов кинофильмов в соответствующем подкорпусе и находит следующие знаки устной речи: введение специальных межфразовых частиц-скреп, эллипсис союзов, экспликация перфоматива, введение метатекста, экспликация ролей, интенсификация фативки, анокопа, рематизация и эмфатизации текста, реализация аннотационного избытка, расхождение между транскриптом и субтитрами и др. [10].

Рассмотрим, какие типы заданий при обучении казахскому языку и литературе на различных ступенях образования может дать поэтический корпус:

1. Поиск слова и его значения в современном казахском языке и в поэзии Абая (сопоставительный анализ).
2. Сопоставление оригинала и перевода с различных позиций (для инофонов).
3. Составление толкового словаря определённого стихотворения (цикла), например, стихотворений о временах года.
4. Анализ возможного изменения семантики слова в тексте Абая и в современных поэтических текстах.
5. Поиск омонимии и разъяснение семантики омонимов.
6. Подбор однокоренных слов на примере поэзии Абая.
7. Синтезирование различных форм одного глагола.
8. Поиск постановки авторских знаков препинания.
9. Поиск авторских неологизмов.
10. Анализ заимствований в поэзии Абая.
11. Этимологический анализ определённого стихотворения, цикла, стихотворений определённого периода.
12. Создание хронологической картины создания стихотворений.
13. Поиск тематически близких стихотворений и написание

14. Поиск примеров слов, употребленных в метафорическом значении.
15. Сравнение употребления слов в поэзии Абая и в современных прозаических/поэтических текстах в корпусе казахского языка (при их наличии) или в других современных источниках.
16. Сопоставление контекстных значений слов в поэтическом корпусе Абая и в современных толковых словарях казахского языка.
17. Поиск в стихотворениях Абая авторских, необычных тропов.
18. Сравнение средств художественной выразительности, трансформирующихся в переводах Абая.
19. Анализ синтаксических конструкций, характерных для поэтического стиля Абая.
20. Выявление диахронического аспекта формирования языковых явлений казахского языка сквозь призму поэзии Абая.
21. Определение эквивалентности переводов текстов Абая на европейские языки, на тюркские языки.
22. Анализ семиотической структуры поэтического корпуса Абая (в жанровом, тематическом и др. видах концентрации).

Одним из важных моментов, отражающих возможности национального корпуса языка, в том числе – казахского, в том числе, поэтического, становится широкое применение возможностей языкового корпуса при профессиональной подготовке будущих учителей-словесников. Все возможные пути использования корпусной лексики, изложенные выше, также с успехом могут быть использованы в курсе методики обучения языку и литературе. В частности, при формировании методической компетенции в контексте умения составлять задания на основе ресурсов национального корпуса.

К принципам, которым необходимо следовать студентам – будущим филологам при составлении заданий на основе корпуса можно отнести следующие:

1. Учет возрастных особенностей учеников, являющихся целевой аудиторией заданий. Необходимо дать понять будущим учителям, что материал, содержащийся в корпусах, не имеет возрастной маркировки и потому при выборе текстов необходим ответственный избирательный подход к лексике.
2. Принцип развивающего обучения
3. Дифференцированный подход или персонализированная позиция в политике составления заданий. Если учебный материал в бумажных учебниках, как правило, рассчитан на среднестатистического ученика, то использование обширного материала национальных корпусов предоставляет широкие возможности для индивидуального подхода.
4. Оптимизация объема материал, используемого для заданий с учетом учебного времени, запланировано затрачиваемого на выполнение.
5. Разнообразие используемого материала – включение, к примеру, мультимедийного материала.
6. Учет реального речевого контекста при рассмотрении языкового явления. Этот принцип означает ситуативный подход к подбору материалов, создание коммуникативно ориентированных диалогических и монологических проблемных моментов.
7. Учет актуальности отбираемого языкового материала. Имеется ввиду востребованность для целевой аудитории «здесь и сейчас».
8. Критическое отношение к отбираемому материалу, использование дополнительных источников лингвистических и литературоведческих знаний.
9. Использование не только нормированных текстов, но и текстов с нарушением норм.
10. Демонстрация активных процессов в языке – изображение языка как развивающегося явления.

Анализ статей по теме позволяет сделать вывод о том, что для составления заданий на основе национальных языковых корпусов может быть реализовано двумя способами: с уже разработанным заданием, когда необходим лишь подходящий материал и для коррекции готовых задач.

Следует отметить также, что использование возможностей национальных языковых корпусов в учебных и методических целях имеет ряд преимуществ, к которым можно отнести экономию времени, повышение мотивации, синхронизацию современной образовательной практики с запросами современности и т.д.

Представленный в материале обзор и исследовательская позиция на сегодняшний день может вызвать научную полемику в том плане, что лингвистическое сообщество разделяется на сторонников и противников корпусных разработок и их использования в обучении и исследованиях. Так, к примеру, Апресян считает, что безоглядное использование ресурсов корпусов может привести к фальсификации вследствие неверного использования количественных данных. Эта позиция, безусловно, заслуживает внимания, однако не исключает ресурсное использование материалов корпусов. Widdowson, H. G. [11] также выступает против безоглядного доверия к материалам корпусов. Соглашаясь с авторами этих идей в том, что в текстах корпусов могут быть количественные, качественные и иные искажения, тем не менее, считаем нужным отстаивать позицию ценности корпусов в исследовательских и методических целях. При этом важным является критический подход к использованию ресурсов.

Заслуживает внимание позиция С.В.Гусаренко и М.К.Гусаренко [12], которые рассматривают возможность создания корпусов текстов с нарушениями языковых и речевых норм для использования в различных, в том числе, в методических целях. При этом, по мнению лингвистов, аккумуляция текстов с ошибками, к примеру, в речи билингвов, может стать ценным материалом и своеобразным «антиобразцом» для всех, кто знакомится с нормированной устной и письменной речью.

Находясь в позиции принятия различных точек зрения позволим себе высказать свою позицию относительно того, что создание поэтического корпуса Абая находится все-таки вне упомянутых рисков, поскольку должен представлять собой образцы классической речи и текстов, взятых из практически канонизированных источников.

Заключение. Под корпусом понимается оцифрованный контент, содержащий все тексты на определенном языке в различном формате и лингвистические размеченные. Рассмотренные пути использования возможностей национальных языковых корпусов в исследованиях и образовании позволяют сделать ряд выводов.

Использование национальных языковых корпусов в образовательном процессе играет значительную роль в повышении качества обучения за счет внедрения информационных технологий. Национальные языковые корпуса не только служат сборниками текстов, но и содержат целый ряд инструментов, расширяющих возможности традиционных учебных материалов.

Важным свойством национальных корпусов является его структурирование по подкорпусам, которое и дает ценный материал для преподавателей и студентов в процессе обучения. Важное значение имеет снятая омонимия.

Национальные корпуса предоставляют следующие возможности:

1. Автоматизированный поиск.

Наличие поисковых систем в корпусах позволяет быстро и эффективно находить конкретные слова, фразы или конструкции в огромном количестве текстов. Это существенно упрощает процесс изучения языковых норм и правил, а также помогает углубленно исследовать особенности языка.

Наличие поисковых систем в корпусах текстов оказывает значительную помощь при изучении языковых норм и правил, а также при исследовании особенностей языка.

Поисковые системы предоставляют возможности для быстрого и эффективного поиска конкретных слова, фразы или конструкции в огромном объеме текстов.

Преимущество корпусного подхода состоит в том, что он позволяет не только найти необходимую лингвистическую и литературоведческую информацию, но и проанализировать ее в контексте, к примеру, дает картину того, как именно этот словоформа употребляется в различных контекстах или в разных стилях речи.

Поисковые системы в корпусах текстов демонстрируют особенности грамматической структуры и контекстуальных значений слов, к примеру, этому служит поиск по коллокациям для определенного слова или фразы.

2. Лингвистические метки.

Лингвистические метки в корпусах транслируют информацию о падежах, временах, частях речи и других грамматических аспектах, используя которые можно автоматизировать морфологическую информацию. Кроме того, лингвистические метки могут указывать на синтаксическую роль слова в предложении, например, подлежащее, сказуемое, дополнение и т.д. Здесь, несомненно, возникает проблема потери навыка самостоятельного определения грамматических категорий. Но эта проблема решается индивидуально в каждом методическом случае.

Корпусная информация позволяет лучше разбираться в структуре языка и более точно анализировать тексты. Эти метки могут использоваться для изучения и понимания грамматических правил и норм, а также для извлечения конкретной информации из текста. Так, корпуса предоставляют возможности для изучения падежей, склонений и согласования слов при изучении родного и иностранных языков.

Кроме того, лингвистические метки в языковых корпусах облегчают работу исследователей и разработчиков лингвистических систем и инструментов. Они позволяют автоматизировать обработку текстов, например, для создания автоматических морфологических анализаторов или синтаксических парсеров.

Лингвистические метки в языковых корпусах являются важным инструментом для изучения языка и проведения лингвистических исследований. Они помогают разбираться в грамматических характеристиках слов и их синтаксических функциях, что способствует более глубокому пониманию и использованию языковых норм и правил. Соответственно, поэтические разметки продвигают литературоведческие изыскания.

3. Нормативные данные.

Национальные языковые корпуса включают в себя словари и справочники, содержащие актуальные нормы языка. Это дает возможность быстро и легко проверить правильность написания, произношения и употребления слов, а также получить их толкования.

Кроме того, национальные языковые корпуса могут также содержать образцы текстов различных жанров и стилей, что позволяет изучать особенности языка в контексте его использования. Это ценный ресурс не только для лингвистов и филологов, но и для всех, кто интересуется языком и его правильным использованием. Национальные языковые корпуса активно используются в образовательных целях, помогая изучать и анализировать язык на основе реальных данных. Этот инструмент способствует развитию лингвистического анализа и понимания языковых явлений.

Благодаря постоянному обновлению и пополнению информации национальные языковые корпуса остаются актуальными и релевантными для исследований в области языкознания, литературоведения и обучения языкам. Это важный инструмент поддержания и стандартизации языка в современном обществе, способствующий его сохранению и развитию.

4. Аутентичность текстов.

Отличительной особенностью языковых корпусов является наличие текстов реального использования языка, максимально приближенных по составу к лексике,

встречающейся в повседневной жизни. В отличие от традиционных учебных текстов, которые могут быть искусственно составлены или адаптированы, тексты в корпусах отражают естественное употребление языка носителями.

Это делает языковые корпуса более реалистичными и полезными для изучения языка, так как они предоставляют учащимся реальные примеры использования слов и выражений в различных контекстах. Благодаря этому, учащиеся могут лучше понять, как использовать язык в реальных ситуациях и развивать свои коммуникативные навыки.

Кроме того, языковые корпуса предоставляют возможность исследования различных аспектов языка, таких как лексика, грамматика, стилистика и семантика, на основе большого объема данных. Это помогает лингвистам и исследователям получить более глубокое понимание языка и его использования в различных контекстах.

5. Корпусный анализ.

Возможность проводить корпусный анализ позволяет изучать закономерности языка, выявлять частоту употребления слов и выражений, проследить семантические связи и исследовать историческую эволюцию языка. Это углубляет понимание языковых процессов и способствует развитию исследовательских навыков.

Важной составляющей корпусного анализа является возможность изучения контекста употребления слов и выражений. Это позволяет лучше понять их значения и различия в использовании в разных ситуациях. Кроме того, анализ семантических связей между словами позволяет выявить тонкие нюансы значения, что полезно как для изучения языка, так и для его перевода и интерпретации.

Корпусный анализ также открывает возможности для изучения изменений в диахроническом и синхроническом аспектах. Анализ частоты употребления слов и выражений в различных временных периодах позволяет наблюдать эволюцию языка и выявлять тенденции его развития. Это особенно важно для лингвистических исследований и создания словарей и грамматик, отражающих современное состояние языка.

В целом, корпусный анализ является мощным инструментом для изучения языка на более глубоком уровне и понимания его функционирования в реальных условиях общения. Он помогает не только лингвистам и переводчикам, но и обычным пользователям языка лучше освоить его структуру и особенности.

6. Лингвистические и литературоведческие исследовательские инструменты.

Языковые корпуса предоставляют доступ к различным лингвистическим и литературоведческим исследовательским инструментам, таким как выявление коллокаций, анализ семантических полей, поиск синонимов и антонимов. Эти инструменты помогают студентам развивать критический анализ и навыки интерпретации языковых данных.

7. База для разработки учебных материалов.

Языковые корпуса служат также источником для разработки современных учебных материалов. Аутентичные тексты, нормативные данные и лингвистические метки корпусов могут быть интегрированы в учебные материалы, обеспечивая более глубокое и всестороннее изучение языка. Использование языкового корпуса становится важным моментом погружения в языковую среду и дает возможность разрабатывать задания как исследовательского, так прагматико-ориентированного активного характера с целью развития устной и письменной речи.

Языковые корпуса являются многосторонним ресурсом для лингвистических, литературоведческих исследований и обучения языку. Они предоставляют доступ к разнообразным инструментам, таким как выявление коллокаций, которые позволяют определить типичные сочетания слов в тексте. Анализ семантических полей позволяет понять смысловые отношения между словами и выявить их значения в контексте. Поиск синонимов и антонимов помогает расширить словарный запас и лучше понять различия в значениях слов. Эти инструменты способствуют развитию критического анализа и

интерпретации языковых данных, что является ключевым навыком для студентов исследователей в области языка и лингвистики.

Хотелось бы отметить еще один актуальный момент, связанный с образовательным контекстом использования языковых корпусов и поэтического корпуса Абая, в частности. Этот аспект связан с формированием функциональной грамотности обучающихся. Речь идет о смысловом чтении художественных текстов. Работа со значениями слов, содержащихся в текстах, поиск семантики и этимологии, поиск историзмов и архаизмов, определение связи с современными значениями слов и другие виды работ сопряжены с осознанным восприятием текстов, проникновением в их структурные и содержательные особенности.

Таким образом, можно сделать вывод о том, что при всех возможных неточностях данных, неполноте сведений или еще каких бы то ни было негативных моментах национальные корпуса являются ценным источником языкового материала и могут быть использованы как в исследовательских, так и методических целях. При этом, несомненно, на пользователей профессионального уровня и на разработчиков ложится большая ответственность за использование корпусных данных в разумных пределах и на основе критического профессионального подхода.

Исследование, в рамках которого написана данная статья, находится на этапе разработки, поэтому многие из представленных здесь идей имеют гипотетический характер. Перспективами исследования является разработка поэтического корпуса Абая Кунанбаева и поиск возможных путей использования его ресурсов в исследовательских, образовательных и методических целях, что должно быть апробировано в ходе экспериментальной работы.

В современной образовательной ситуации языковой корпус уже признан одним из удобных инструментов работы с текстом. К положительным сторонам использования поэтического корпуса Абая можно отнести возможность лингвистического и литературоведческого анализа, развития устной и письменной речи, возможность верификации знаний, помощь различного рода разметки, позволяющей моделировать и осуществлять поиск по различным заданным параметрам, в зависимости от учебной ситуации.

Список использованных источников:

1. Захаров В.П. Поисковые системы Интернета как инструмент лингвистических исследований // Русский язык в интернете. Казань, 2003. С. 48-59.
2. Воробей И.А. Конкорданс в составе метода анализа поэтического концепта. <https://www.academia.edu/>
3. Конкорданс. Словарная статья <http://www.lomonosov-fund.ru/enc/ru/encyclopedia:0127200#googlevignette>
4. Орехов Б.В. Еще раз об исследовательском потенциале поэтического корпуса: метр, лексика, формула. <https://publications.hse.ru/articles/178496831>
5. Гаспаров М. Л. Метр и смысл. Об одном из механизмов культурной памяти. — М.: Фортуна ЭЛ, 2012. — С.3-16.
6. Дерябина И.В. Характеристика педагогической работы с английским национальным корпусом. Гуманитарные науки. Педагогика и психология. ISSN 1810-0201. Вестник ТГУ, выпуск 9 (113), 2012. — С. 156-160
7. Рыков В.В. Прагматически ориентированный корпус текстов // Тверской лингвистический меридиан. Тверь, 1999. Вып. 3. С. 89-96.
8. Добрушина Р.М. Как использовать национальный корпус русского языка в образовании. <https://ruscorpora.ru/old/sbornik2005/18dobrushina.pdf>
9. Анохина С.А. Использование материалов национального корпуса русского языка при изучении лексических категорий// Национальный корпус русского языка и

- проблемы гуманитарного образования. Под редакцией Н.Р. Добрушиной. – Москва 2007. С. 150-159. \
10. Е.А.Гришина О маркерах разговорной речи (предварительное исследование подкорпуса кино в национальном корпусе русского языка, www.ruscorpora.ru)
 11. Widdowson, H. G. (2000). On the limitations of applied linguistics. Applied Linguistics, 21, 3-25. <https://doi.org/10.1093/applin/21.1.3>
 12. С.В.Гусаренко, М.К.Гусаренко О корпусе речевых образцов с ошибками в употреблении русского языка как иностранного: способы представления данных и параметры глубинной разметки//Гуманитарные и юридические исследования. 2022. Т. 9 (4)

SEMANTIC AMBIGUITY PROCESSING IN HUMAN COGNITION: A COMPARATIVE ANALYSIS OF NLP MODELS AND PSYCHOLINGUISTIC APPROACHES

Suyunova Mohinur Ilxom qizi

PhD student, National University of Uzbekistan

Abstract: Semantic ambiguity is a pervasive phenomenon in both natural language use and computational language processing. This paper investigates how semantic ambiguities are processed in human cognition from a psycholinguistic perspective and compares these processes to the mechanisms employed by modern Natural Language Processing (NLP) models. Drawing on insights from cognitive science, linguistics, and artificial intelligence, we explore how humans resolve ambiguity dynamically using context, prior knowledge, and cognitive heuristics, while machine learning models rely primarily on statistical patterns and pre-trained embeddings. We also discuss theoretical frameworks such as constraint-based models and modularity of mind to better understand the differences between human and machine ambiguity resolution. Finally, we consider how integrating psycholinguistic theories into NLP design could advance the development of more cognitively plausible language technologies.

Keywords: Semantic ambiguity, human cognition, NLP models, psycholinguistics, language processing, context, disambiguation, constraint-based models, modularity, incremental processing.

Аннотация: Семантическая неопределенность — это широко распространённое явление как в естественном использовании языка, так и в области вычислительной обработки текста. В данной статье исследуется, как семантическая неопределенность обрабатывается в человеческом сознании с точки зрения психолингвистики, и проводится сравнение этих процессов с механизмами, используемыми современными моделями обработки естественного языка (NLP). Основываясь на знаниях из когнитивной науки, лингвистики и искусственного интеллекта, мы изучаем, как человек динамически разрешает неопределенность, используя контекст, предшествующие знания и когнитивные эвристики, в то время как модели машинного обучения полагаются в основном на статистические закономерности и заранее обученные эмбединги. Также рассматриваются теоретические концепции, такие как модели на основе ограничений и модульность разума, для более глубокого понимания различий между человеческими и машинными процессами разрешения неопределенности. Наконец, обсуждаются перспективы интеграции психолингвистических теорий в разработку NLP-систем для создания более когнитивно достоверных языковых технологий.

Ключевые слова: Семантическая неопределенность, человеческое сознание, модели NLP, психолингвистика, обработка языка, контекст, разрешение неопределенности, модели на основе ограничений, модульность, инкрементальная обработка

Introduction

Semantic ambiguity arises when a linguistic unit (such as a word, phrase, or sentence) allows multiple interpretations. Understanding how humans resolve these ambiguities offers valuable insights for building more sophisticated and human-like NLP systems. Psycholinguistic studies reveal that ambiguity resolution in human cognition is an active, context-driven process influenced by multiple cognitive mechanisms. In contrast, current NLP models often depend on learned statistical patterns. This paper aims to bridge these perspectives by analyzing commonalities and divergences in human and machine semantic processing, while grounding our discussion in core psycholinguistic theories.

The Nature of Semantic Ambiguity

Semantic ambiguities can be classified as:

- Lexical ambiguity (e.g., 'bank' as financial institution or river edge)
- Syntactic ambiguity (e.g., 'Visiting relatives can be boring')
- Pragmatic ambiguity (e.g., ironic statements)

Resolution of ambiguity relies heavily on context, prior discourse, syntactic structure, and pragmatic cues. Psycholinguistic models emphasize that ambiguity triggers multiple simultaneous interpretations that compete until one meaning becomes dominant.

Psycholinguistic Theories on Ambiguity Processing

Psycholinguistic research provides several key frameworks for understanding ambiguity resolution:

Constraint-Based Models: Constraint-based models argue that multiple sources of information (syntax, semantics, pragmatics, frequency) are used simultaneously to constrain and resolve interpretations.

Modularity of Mind: Fodor's modularity hypothesis suggests linguistic processing occurs within a domain-specific system.

Garden-Path Theory: Garden-path theory proposes humans initially opt for the simplest syntactic structure, often necessitating reanalysis.

Predictive Processing Frameworks: Predictive processing frameworks posit continuous prediction of upcoming input to resolve ambiguity.

Experimental methods such as eye-tracking, ERP, and fMRI support the view that ambiguity resolution is dynamic, involving activation, competition, and selection among alternatives.

NLP Models and Semantic Ambiguity

Modern NLP models like BERT, GPT-3, and T5 encode contextual dependencies to address ambiguity. However:

- They lack genuine inferential pragmatics.
- Context handling is limited by training data.
- Errors occur in novel or less frequent contexts.

Transformer models represent context dynamically but remain primarily statistical rather than truly inferential.

Comparative Analysis of Human and Machine Processing

Humans leverage syntactic cues, semantic plausibility, and real-world expectations. NLP models prioritize surface-level statistical patterns. Key differences include:

Context utilization: Deeply inferential vs. statistical co-occurrence.

Ambiguity handling: Heuristic-driven vs. pattern recognition.

Learning method: Continuous experience vs. static pre-training.

Error patterns: Garden-path effects vs. out-of-distribution errors.

Adaptability: Rapid adaptation vs. brittle fine-tuning.

Toward Cognitive-Inspired NLP Models

Future NLP models could incorporate:

- Cognitive heuristics for real-time disambiguation.
- Incremental parsing architectures.
- Pragmatic inference modules.
- Real-time context learning mechanisms.

Modeling human strategies may improve robustness, interpretability, and human-likeness.

Conclusion

Semantic ambiguity challenges both humans and machines, but mechanisms differ profoundly. Human processing is context-rich and heuristic, while NLP models are statistical. Future language technologies can advance by drawing on psycholinguistic theories and cognitive architectures.

References:

1. Altmann, G. T. M. (1998). Ambiguity in Sentence Processing. Trends in Cognitive Sciences, 2(5), 146–152.
2. MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical Nature of Syntactic Ambiguity Resolution. Psychological Review, 101(4), 676–703.
3. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL-HLT.
4. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. In Advances in Neural Information Processing Systems.
5. Ferreira, F., & Clifton, C. (1986). The Independence of Syntactic Processing. Journal of Memory and Language, 25(3), 348–368.
6. Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? Language, Cognition and Neuroscience, 31(1), 32–59.
7. Fodor, J. A. (1983). The Modularity of Mind: An Essay on Faculty Psychology. MIT Press.
8. Abdurakhmonova, N. Z., & Teshabayev, A. G. (2019). Modeling IT of discourse analysis and the issues machine translation. Scientific Bulletin. Physical and Mathematical Research, 1(2), 101-107.

MACHINE LEARNING MODELLARI ASOSIDA O‘ZBEK TILIDAGI AUDIOMATNLARNI TAHLIL QILISH: KUN.UZ MISOLIDA

Urazaliyeva Mavluda Yangiboyevna

O‘zbekiston Milliy universiteti mustaqil izlanuvchisi

E-mail: urazaliyeva_m@nuu.uz

Annotatsiya: Mazkur maqolada o‘zbek tilidagi audiomatnlarni avtomatik tarzda transkripsiyalovchi mashinaviy o‘rganish modellarining fonetik va grammatik aniqlik darajasi chuqur ilmiy tahlil qilinadi. Tadqiqotda Whisper, wav2vec 2.0, CTC va Seq2Seq modellarining ish faoliyati og‘zaki nutq namunalari — kun.uz saytining 2023–2024-yillardagi audio yangiliklari asosida qiyoslash tarzda baholandi. Har bir modelning Word Error Rate (WER) va Character Error Rate (CER) mezonlari bo‘yicha samaradorligi o‘zbek tilining fonetik murakkabligi, morfologik affikslar tizimi va urg‘u o‘zgarishlari nuqtai nazaridan tekshirildi. Olingan natijalar tahlil qilinib, modellar orasidagi afzalliklar hamda ularning zaif jihatlari ko‘rsatib berildi. Shuningdek, maqolada xalqaro ASR korpuslari bilan taqqoslash asosida o‘zbek tiliga moslashtirilgan modellar yaratishning dolzarbligi ilmiy asoslangan xulosalar bilan yoritildi.

Kalit so‘zlar: audiomatn, avtomatik transkripsiya, fonetik xususiyatlar, morfemik tahlil, ASR modellari, WER, CER.

Аннотация: В данной статье представлен глубокий научный анализ фонетической и грамматической точности моделей машинного обучения, автоматически транскрибирующих аудиотексты на узбекском языке. В исследовании проводилась сравнительная оценка эффективности моделей Whisper, wav2vec 2.0, CTC и Seq2Seq на основе образцов устной речи – аудио новостей сайта kun.uz за 2023-2024 годы. Эффективность каждой модели оценивалась по критериям Word Error Rate (WER) и Character Error Rate (CER) с учетом фонетической сложности узбекского языка, его системы морфологических аффиксов и вариаций ударения. Полученные результаты были проанализированы, выделены преимущества и недостатки каждой модели. Кроме того, в статье подчеркивается актуальность разработки моделей, адаптированных к узбекскому языку, на основе сравнений с международными корпусами ASR, с научно обоснованными выводами.

Ключевые слова: аудио текст, автоматическая транскрипция, фонетические особенности, морфемный анализ, модели ASR, WER, CER.

Abstract: This article provides an in-depth scientific analysis of the phonetic and grammatical accuracy of machine learning models that automatically transcribe audio texts in the Uzbek language. The study comparatively evaluated the performance of Whisper, wav2vec 2.0, CTC, and Seq2Seq models based on oral speech samples - audio news from the kun.uz website for 2023-2024. The effectiveness of each model was assessed according to Word Error Rate (WER) and Character Error Rate (CER) criteria, taking into account the phonetic complexity of the Uzbek language, its system of morphological affixes, and stress variations. The obtained results were analyzed, highlighting the advantages and weaknesses of each model. Furthermore, the article emphasizes the relevance of developing models adapted to the Uzbek language, based on comparisons with international ASR corpora, supported by scientifically grounded conclusions.

Keywords: audio text, automatic transcription, phonetic features, morphemic analysis, ASR models, WER, CER.

Kirish

Ma'lumki, avtomatik nutqni matnga aylantirish tizimlari (ASR — Automatic Speech Recognition) sun'iy intellektning eng faol rivojlanayotgan yo'nalishlaridan biridir [Jurafsky &

Martin., 2020: 312]. Bunday tizimlar turli tillarda, ayniqsa ingliz, nemis, ispan tillarida katta muvaffaqiyatga erishgan bo'lsa-da, agglutinatív tuzilishga ega o'zbek tilida hali yetarlicha aniqlik va moslashuv darajasiga ega emas [Xolmatova va boshq., 2022: 123]. Aynan shuning uchun ham, fonetik va morfemik murakkabliklarga ega o'zbek tilida audiomatnlar bilan ishlashda machine learning modellarining imkoniyatlarini chuqur o'rganish zarurati tug'iladi.

Shu bois, ushbu maqolada mashinaviy o'rganish modellarining o'zbek tilidagi audiomatnlar ustida fonetik va grammatik aniqlikka erishish darajasi kun.uz saytidagi misollar orqali baholanadi. ASR modellarining fonetik o'xshashlik, talaffuzdagi qisqarish, urg'ular va morfemalarning segmentatsiyasi nuqtai nazaridan qanday ishlashini aniqlash maqsad qilib olingan.

Uslubiyat

Ushbu tadqiqot doirasida qo'llanilgan metodologik yondashuvlar mazmunan ilmiy aniqlik, empirik dalillar asosida tanlab olingan ma'lumotlar bazasi va mashinaviy o'rganish modellarining xolis bahosiga tayangan. Tadqiqot bir necha bosqichlarda amalga oshirildi: birinchi bosqichda, kun.uz saytining 2023–2024 yillardagi rasmiy audio xabarlar asosida nutq matnlari tanlab olindi. Har bir audiomatnga mos matnli versiya (subtitle) mavjud bo'lib, bu avtomatik transkripsiyalarning aniqligini baholash imkonini berdi.

Ikkinchi bosqichda, to'rt xil mashinaviy o'rganish modeli – Whisper, wav2vec 2.0, CTC va Seq2Seq – sinovdan o'tkazildi. Bu modellar tanlanishida ularning texnologik asoslari (transformer, self-supervised, attention, CTC algoritmlari) va turli tillarda testdan o'tkazilgan tajribalardagi samaradorlik ko'rsatkichlari asosiy mezon sifatida qaraldi [Radford et al., 2022: 9; Baevski et al., 2020: 6; Graves et al., 2006: 369; Bahdanau et al., 2015: 4].

Uchinchi bosqichda esa, transkripsiya natijalari an'anaviy WER va CER ko'rsatkichlari yordamida tahlil qilindi. Har bir modelning xatolik darajasi, ayniqsa o'zbek tiliga xos urg'ular, fonetik qisqarishlar, affiksalar murakkabliklar va segmentatsion zaifliklar nuqtai nazaridan chuqur tahlil qilindi. Bu bosqichda har bir model uchun natijalar grafik va jadval asosida tasvirlandi.

Ushbu uslubiyat aynan amaliy korpusga tayanishi, mashinaviy modellarni real nutq muhitida sinovdan o'tkazishi bilan qadrlidir. Bu yondashuv kelgusida o'zbek tilida maxsus ASR modellari yaratishda asosiy poydevor vazifasini bajaradi..

Natijalar

Model sinovlari natijasida turli algoritmlar asosida ishlaydigan ASR tizimlarining o'zbek tilidagi audiomatnlarni qayta ishlashdagi imkoniyatlari aniqlandi. Baholash mezonlari sifatida qabul qilingan WER (so'z xatoligi ko'rsatkichi) va CER (belgi xatoligi ko'rsatkichi) qiymatlari asosida har bir modelning fonetik va morfemik tahlildagi samaradorligi taqqoslandi [Radford et al., 2022: 9; Baevski et al., 2020: 5; Graves et al., 2006: 376].

Aniqlangan natijalarga ko'ra, Whisper modeli eng past WER (11.2%) va CER (6.7%) ko'rsatkichlari bilan yetakchilik qildi. Ushbu model ko'p tilli ma'lumotlar bazasi asosida o'qitilgan bo'lib, fonetik qisqarishlar, urg'u o'zgarishlari va affiksalar murakkabliklarni sezish darajasi nisbatan yuqori bo'lib chiqdi. Masalan, “tashrif buyurdi” iborasi turli kontekstlarda qisqargan holda aytilgan bo'lsa ham, model bu segmentni to'g'ri ajrata oldi.

Wav2vec 2.0 modelining ishlashi fonetik jihatdan ancha barqaror bo'lib (WER: 18.4%), morfemik segmentatsiyada sustlik kuzatildi. Ayniqsa, fe'l va affiks o'rtasidagi chegarani aniqlashda “ketmoqda” va “ketayapti” kabi iboralarda xatolar ko'p kuzatildi. Natijalarda substitution va deletion xatolari yuqori bo'ldi.

CTC modelida esa xatolik ko'rsatkichlari eng yuqori (WER: 24.9%, CER: 13.5%) bo'lib, bu uning vaqtga asoslangan segmentatsiya metodologiyasi bilan bog'liq. Jumladan, “hukumat bayonoti tinglandi” degan gapda model “hukumat bayoni” shaklida segmentatsiya qilgan. Bu esa morfologik tahlilning noto'g'ri bajarilganini ko'rsatadi [kun.uz 2023–2024].

Seq2Seq modeli esa kontekstni chuqur o'rganishga qaratilgan bo'lsa-da [Bahdanau et al., 2015: 4], og'zaki nutqda tez aytilgan va fonetik jihatdan qisqargan so'z birikmalarini to'g'ri tahlil qilishda muammolarga duch keldi. Misol uchun, “yo'l harakati qoidalari” jumlasini “yulharakat

qoidasi” sifatida transkripsiyalandi [Kun.uz, 2024]. Bu holat CER qiymatining yuqori bo‘lishiga olib keldi.

Modellarni baholash natijalariga ko‘ra, ularning aniqligi arxitektura, til korpusining hajmi va sifati hamda o‘zbek tiliga xos fonetik va morfologik xususiyatlarga moslashuvchanligi bilan bevosita bog‘liq.

O‘tkazilgan eksperimentlar fonida shuni ko‘rish mumkinki, turli mashinaviy o‘rganish modellarining o‘zbek tilidagi audiomatnlarni tahlil qilish imkoniyatlari bir-biridan sezilarli darajada farq qiladi. Jumladan, Whisper modeli fonetik qisqarishlar va urg‘ularni sezish hamda ularni kontekstga mos holda transkripsiyalashda yuqori samaradorlikni ko‘rsatdi. Bu holat modelning ko‘p tillilikka moslashtirilgani va o‘qitish hajmining kengligi bilan bog‘liq.

Wav2vec 2.0 modeli esa fonetik jihatdan sezilarli darajada yaxshi ishlagan bo‘lsa-da, affikslar va grammatik birliklar o‘rtasidagi chegaralarni aniqlashda muammolarga duch keldi. Masalan, “olib kelmoqda” iborasining “olibkelmoq” shaklida bir butun sifatida qabul qilinishi morfemik tahlil xatolariga olib keldi.

CTC modelida segmentatsiya vaqtga bog‘langanligi bois morfemik chegaralar ko‘pincha noto‘g‘ri aniqlangan. Bu ayniqsa, birikma affikslarda, ya‘ni “-yapti”, “-moqda” kabi fe‘l formalari bilan yakunlanadigan gaplarda sezilarli bo‘lgan.

Seq2Seq modelida esa kontekstni tushunishga moyillik yuqori bo‘lsa-da, nutqdagi fonetik qisqarishlar, shoshilinch talaffuz va urg‘u farqlari sababli ayrim transkripsiya natijalari noto‘g‘ri chiqdi. Jumladan, “hisobot taqdim etildi” iborasi ba‘zida “isobottaqdimetdi” shaklida birlashtirilib yozilgan [Kun.uz, 2024].

Umuman olganda, modellar orasida Whisper yuqori aniqlik bilan ajralib turgan bo‘lsada, har bir modelning o‘ziga xos kuchli va zaif tomonlari mavjud. Ushbu natijalar o‘zbek tiliga moslashtirilgan, agglutinativ tuzilmani inobatga oluvchi maxsus ASR modellarini ishlab chiqish zarurligini ta’kidlaydi.

Xulosa

Ushbu tadqiqot doirasida olib borilgan tajribalar mashinaviy o‘rganish modellarining o‘zbek tilidagi audiomatnlar bilan ishlashdagi samaradorligini chuqur tahlil qilish imkonini berdi. Jumladan, Whisper modeli WER va CER ko‘rsatkichlari bo‘yicha eng yuqori aniqlik darajasini namoyon etdi. Modellar fonetik qisqarishlar, urg‘u tizimi, affiks morfologiya kabi jihatlar bilan bog‘liq xatolarni har xil darajada qayd etgani sababli ularning ishlash samaradorligi sezilarli darajada farq qildi.

Wav2vec 2.0 modeli fonetik sathdagi aniqlikka ega bo‘lsa-da, morfemik segmentatsiyada ayrim cheklovlar ega ekanligi aniqlandi (Baevski et al., 2020). Ayniqsa, agglutinativ tuzilmadagi affikslarning ko‘p shaklli namoyon bo‘lishi bu modelda transkripsiya aniqligiga salbiy ta’sir ko‘rsatdi.

Shuningdek, CTC va Seq2Seq modellarida segmentatsiya va kontekstni anglashdagi cheklovlar sababli xatoliklar soni yuqori bo‘ldi. CTC modelining vaqt asosidagi segmentatsiyasi fonetik jihatdan noaniq bo‘lib, “tashrif buyurdi” yoki “bayonot taqdim etildi” kabi iboralarda affikslar noto‘g‘ri ajratilgan. Seq2Seq modelida esa kontekstga qarab noto‘g‘ri fonetik segmentlar paydo bo‘lishi kuzatildi.

Yakunda shuni ta’kidlash joizki, ushbu tadqiqot orqali o‘zbek tilida audiomatnlar bilan ishlovchi maxsuslashtirilgan, fonetik va morfologik darajada optimallashtirilgan ASR modellarini ishlab chiqish zarurati isbotlandi.

Foydalanilgan adabiyotlar:

1. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. OpenAI. <https://openai.com/research/whisper>

2. Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. arXiv preprint arXiv:2006.11477. <https://arxiv.org/abs/2006.11477>
3. Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006). Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. Proceedings of the 23rd International Conference on Machine Learning (ICML), 369–376. <https://doi.org/10.1145/1143844.1143891>
4. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies, 10(8), 59-66.
5. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. ICLR. <https://arxiv.org/abs/1409.0473>
6. Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed.). New Jersey: Prentice Hall. ISBN: 978-0131873216
7. Халматова, Н., Саидов, У., Мирзарахимова, З. (2022). Агглютинатив тилларда морфологик таҳлилнинг муаммолари. Тошкент: Фан нашриёти.
8. Kun.uz (2023–2024). O‘zbekistonda rasmiy axborot yangiliklari va audio-matnlar arxivi. Toshkent: Kun.uz axborot xizmati. <https://kun.uz>
9. Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). Librispeech: An ASR corpus based on public domain audio books. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5206–5210. <https://doi.org/10.1109/ICASSP.2015.7178964>
10. Rousseau, A., Deléglise, P., & Estève, Y. (2014). Enhancing the TED-LIUM corpus with selected data for language modeling and more TED talks. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 3935–3939.
11. Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., & Dahlgren, N. L. (1993). DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM. Linguistic Data Consortium. <https://catalog.ldc.upenn.edu/LDC93S1>
12. Kubedinova L. Khusainov A., Suleymanov D., Gilmullin R., Abdurakhmonova N. First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems. Proceedings of the Computational Models in Language and Speech Workshop (CMLS 2020) co-located with 16th International Conference on Computational and Cognitive Linguistics (TEL 2020) .2020/11: 90-101.
13. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. In Proceedings of the International Conference on Language Technologies for All (LT4All) 2019.
14. N. Abdurakhmonova, U. Tuliyeu and A. Gatiatullin, "Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus.uz," 2021 International Conference on Information Science and Communications Technologies (ICISCT), 2021, pp. 1-4, <http://10.1109/ICISCT52966.2021.9670043>

MILLIY KONTENT YARATISHDA ADABIY TIL ME'YORLARIGA RIOYA QILISH

Tursunboyeva Sitorabonu

O'zbekiston milliy universiteti
Jurnalistika va o'zbek filologiyasi fakulteti
1-kurs tayanch doktoranti

Annotatsiya: Ushbu ilmiy maqola milliy kontent yaratishda adabiy til me'yorlariga rioya qilishning ahamiyatini o'rganadi. Adabiy til me'yorlari milliy identifikatsiyaning til orqali ifodalangan qat'iy qoidalarini va madaniy qiymatlarini belgilaydi. Milliy kontent yaratishda adabiy til me'yorlariga rioya qilish, tilni to'g'ri, samarali va madaniy jihatdan to'g'ri ishlatish imkonini beradi. Maqola, shuningdek, milliy kontentda adabiy til me'yorlarining qanday tatbiq etilishi, ularning tilshunoslik, madaniyatshunoslik va ijtimoiy-psixologik jihatlar bilan bog'liq aspektlari hamda kontent yaratishda to'g'ri til me'yorlarining o'rni haqida tahlilni taklif etadi.

Kalit so'zlar: adabiy til, me'yorlar, milliy kontent, tilshunoslik, madaniyat, til me'yorlari, ijtimoiy-psixologik tahlil.

Abstract: This scientific article examines the significance of adhering to literary language norms when creating national content. Literary language norms define the strict rules and cultural values expressed through language that shape national identity. By following these norms, national content creation can ensure the proper, effective, and culturally appropriate use of language. The article also explores the application of literary language norms in national content, their connection with linguistics, cultural studies, and social-psychological aspects, and the role of adhering to proper linguistic norms in content creation.

Keywords: literary language, norms, national content, linguistics, culture, linguistic norms, social-psychological analysis.

Аннотация: Данная научная статья рассматривает важность соблюдения литературных языковых норм при создании национального контента. Литературные языковые нормы определяют строгие правила и культурные ценности, выражаемые через язык, которые формируют национальную идентичность. Соблюдение этих норм при создании национального контента позволяет обеспечить правильное, эффективное и культурно адекватное использование языка. Статья также анализирует применение литературных языковых норм в национальном контенте, их связь с лингвистикой, культурологией и социально-психологическими аспектами, а также роль соблюдения правильных языковых норм в процессе создания контента.

Ключевые слова: литературный язык, нормы, национальный контент, лингвистика, культура, языковые нормы, социально-психологический анализ.

Milliy kontent yaratish jarayoni til va madaniyatning o'zaro bog'liqligi asosida shakllanadi. Har bir millatning o'ziga xos madaniy qadriyatlari, an'analari va til xususiyatlari milliy identifikatsiyani ifodalashda muhim o'rin tutadi. Milliy kontentda adabiy til me'yorlariga rioya qilish, nafaqat tilni to'g'ri va samarali ishlatish, balki milliy madaniyatni saqlash va uning ijtimoiy-psixologik o'ziga xosliklarini aks ettirish imkonini beradi. Bu me'yorlar tilni ijtimoiy muhitda, madaniy va urf-odatlariga mos ravishda ishlab chiqilishiga yordam beradi. Shuningdek, adabiy til me'yorlarining milliy kontentdagi roli tilning estetik, kommunikativ va madaniy funksiyalarini amalga oshirishda muhim ahamiyatga ega.

Milliy kontent yaratish jarayonida adabiy til me'yorlariga rioya qilishning ahamiyati katta. Adabiy til me'yorlari tilning to'g'ri, aniq va samimiy ishlatilishini ta'minlaydi, bu esa o'z navbatida milliy identifikatsiyaning kuchayishiga olib keladi. Me'yorlar tilning grammatik va leksik xususiyatlarini belgilab beradi, ularni doimiy ravishda saqlash va rivojlantirish esa milliy

kontentning sifatini oshiradi. Misol uchun, o'zbek tilida adabiy til me'yorlari, yozuv va nutq faoliyatining samarali amalga oshirilishida muhim rol o'ynaydi. Shuningdek, adabiy til me'yorlari madaniy qadriyatlarni til orqali ifodalashda muhim vositadir, chunki ular xalqning ruhiyati va an'analari bilan chambarchas bog'liq.

Adabiy til me'yorlari, tilshunoslikda tilning grammatik, leksik va stilistik jihatlarini tashkil etuvchi qoidalar to'plami sifatida qaraladi. Ular tilni to'g'ri va samarali ishlatish uchun zarur asos bo'lib, nafaqat nutqning shakllanishi, balki madaniy qadriyatlarning va milliy identifikatsiyaning ifodalanuvchi vositasi sifatida ham muhim ahamiyatga ega. Til me'yorlari, har bir millatning o'ziga xos til madaniyati, til tizimi va shakllanish jarayonlarining yakuniy natijasidir.

Adabiy til me'yorlariga rioya qilish tilni mukammallashtirishga yordam beradi, bu esa milliy kontentning sifatini oshirishga olib keladi. Shu bilan birga, me'yorlarga rioya qilish, milliy tilning barcha qatlamlarini, xususan, tilning tarixiy va madaniy aspektlarini saqlab qolishga imkon yaratadi. Milliy til, uning leksikasi va grammatikasi orqali milliy tarixni, qadriyatlarni, urf-odatlarini va ijtimoiy mohiyatni aks ettiradi. Shuning uchun, adabiy til me'yorlariga rioya qilinishi tilning madaniy rolini mustahkamlaydi va milliy identifikatsiyaning davomiyligini ta'minlaydi[1]

Tilning grammatik me'yorlari til tizimining asosini tashkil etadi. Grammatik qoidalar tilning sintaktik va morfologik tuzilishini belgilaydi, bu esa nutqning aniq va izchil bo'lishini ta'minlaydi. Adabiy til me'yorlariga rioya qilish, tilning grammatik asoslarini saqlashga va o'zgarishlar yoki xatoliklarni minimallashtirishga yordam beradi. Bu tilni aniq ifodalashga imkon yaratadi va shuningdek, milliy tilning zamonaviy va tarixiy shakllarini uyg'unlashtiradi.

Leksik me'yorlar esa tilning lug'aviy boyligini belgilaydi. Adabiy til me'yorlariga rioya qilish, tilning to'g'ri so'z boyligi bilan ishlashni ta'minlaydi. Leksikada mavjud bo'lgan atamalar, so'zlar va iboralar, milliy madaniyatning o'ziga xos xususiyatlarini aks ettiradi. Leksik me'yorlarga rioya qilish milliy tilning an'anaviy boyligini saqlash va yangi leksik birikmalarni to'g'ri qo'llash imkonini beradi. Milliy kontentda bu leksik birliklarning samarali ishlatilishi, o'sha millatning madaniy qadriyatlari va tarixiy tafakkurini to'g'ri aks ettirishga xizmat qiladi.

Stilistik me'yorlar esa tilning ifodaviy imkoniyatlarini belgilaydi. Stilistika, tilning turli vaziyatlar uchun to'g'ri ishlatilishini ta'minlaydi va tilga estetik qiymat qo'shadi. Adabiy til me'yorlari, turli janrlar va shakllar uchun mos uslubni tanlashga yordam beradi. Milliy kontent yaratishda, stilistik me'yorlarga rioya qilish, tilning emotsional va psixologik xususiyatlarini to'g'ri aks ettirishga yordam beradi. Shu tariqa, tilning stilistik jihatlarini milliy madaniyatning nozik jihatlarini, xalqning ruhiyatini va qadriyatlarini tasvirlashda muhim rol o'ynaydi[2]

Adabiy til me'yorlariga rioya qilish, tilni faqat grammatik yoki leksik jihatdan to'g'ri ishlatish emas, balki uning madaniy aspektlarini ham muhim jihatlarini o'z ichiga oladi. Bu esa milliy identifikatsiyani yaratish va saqlashda, ijtimoiy va madaniy qadriyatlarni to'g'ri ifodalashda zaruriy shart bo'ladi. Milliy kontentda tilning badiiy, psixologik va madaniy yuksakligini ta'minlashda adabiy til me'yorlari alohida ahamiyatga ega.

Adabiy til me'yorlarini buzish, kontentni beqarorlashtirishi va uning ijtimoiy jihatdan noto'g'ri talqin etilishiga olib kelishi mumkin. Milliy kontent yaratishda tilning ijtimoiy va madaniy aspektlari alohida e'tiborga olinishi zarur. Buning uchun tilshunoslik va madaniyatshunoslik yondashuvlaridan foydalanish, tilni faollashtirishda va milliy madaniyatni ifodalashda muhim vosita sifatida ishlatiladi.

Adabiy til me'yorlarini milliy kontent yaratishda qo'llash bo'yicha metodologik yondashuv, tilshunoslik va madaniyatshunoslik tadqiqotlaridan kelib chiqadi. Ushbu yondashuv tilni faollashtirishda uning madaniy jihatlarini hisobga olishni ta'minlaydi. Adabiy til me'yorlarini milliy kontentda tatbiq qilish uchun tilshunoslar, yozuvchilar va o'qituvchilar tomonidan ishlatiladigan grammatik va leksik qoidalar aniqlanadi. Bundan tashqari, ijtimoiy-psixologik kontekstda, milliy kontentni yaratishda tilning ijtimoiy ahamiyati va madaniy kodlar bilan bog'liqligi tahlil qilinadi. Bu jarayonda interdisipliner yondashuv, til va madaniyatni birgalikda tahlil qilish orqali milliy kontentning samaradorligini oshirishga yordam beradi.

Masalan, o'zbek tilida "mehmondo'stlik" tushunchasi, milliy identifikatsiyaning bir qismi sifatida tilga yuksak ahamiyat beriladi. "Mehmondo'st" so'zi o'zbek xalqi uchun nafaqat bir fe'l, balki milliy qadriyatni aks ettiruvchi muhim madaniy koddir. Bunda adabiy til me'yorlariga rioya qilish bu tushuncha va boshqa milliy qadriyatlarning to'g'ri ifodalashiga yordam beradi. Agar "mehmondo'st" so'zi noto'g'ri yoki kontekstdan tashqarida ishlatilsa, bu milliy kontentni beqarorlashtirishi va ijtimoiy jihatdan noto'g'ri talqin etilishiga olib kelishi mumkin.

Bundan tashqari, adabiy til me'yorlari madaniy qadriyatlarni til orqali ifodalashda muhim vosita sifatida xizmat qiladi. Til faqat kommunikatsiya vositasi emas, balki madaniy, ijtimoiy va psixologik o'ziga xosliklarni aks ettiruvchi elementdir. Masalan, o'zbek tilida ishlatiladigan "yurakdan" (mehribonlik, samimiyat) va "ko'ngil" (ichki dunyo, his-tuyg'ular) kabi leksik birliklar madaniy qadriyatlarning tilga qanday aks etishini ko'rsatadi. Bu kabi iboralar, madaniyatning ijtimoiy va emotsional jihatlarini ifodalashda muhim rol o'ynaydi. Agar bu so'zlar adabiy til me'yorlaridan chetga chiqib ishlatilsa, madaniy qadriyatlar noto'g'ri yoritilishi mumkin, bu esa milliy kontentning noto'g'ri qabul qilinishiga olib keladi[3]

"Mehmondo'stlik" tushunchasi – o'zbek tilida bu so'z ijtimoiy qadriyatlarni ifodalashda muhim rol o'ynaydi. "Mehmondo'st" atamasi, o'zbek xalqi uchun samimiy, ochiq va o'zini hurmat qilgan muomala madaniyatini anglatadi. Agar bu so'z milliy kontentda noto'g'ri ishlatilsa (masalan, xususiy yoki salbiy ma'nolarda), bu milliy identifikatsiyaning noto'g'ri ifodalanishiga olib kelishi mumkin. Adabiy til me'yorlariga rioya qilish bu atamaning aniq ma'nosi va mohiyatini saqlashga yordam beradi.

"Ko'ngil" va "yurakdan" iboralari – bu leksik birliklar o'zbek xalqining ichki dunyosini, hissiyotlarini va mehribonligini aks ettiradi. Ularning noto'g'ri ishlatilishi madaniy kodlarning buzilishiga olib keladi, natijada milliy kontent noto'g'ri va tushunarsiz bo'lishi mumkin. Misol uchun, "ko'ngil" so'zining salbiy kontekstda ishlatilishi o'zbek tilining madaniy va emotsional jihatlarini noto'g'ri aks ettiradi[4]

Milliy kontent yaratishda adabiy til me'yorlariga rioya qilish til va madaniyat o'rtasidagi o'zaro bog'lanishning to'g'ri aks etishini ta'minlaydi. Me'yorlar nafaqat tilning grammatik va leksik tarkibini belgilaydi, balki milliy identifikatsiyaning mustahkamlashiga yordam beradi. Adabiy til me'yorlarining buzilishi esa, kontentni beqarorlashtirib, ijtimoiy jihatdan noto'g'ri talqin etilishiga olib kelishi mumkin. Shuning uchun, milliy kontent yaratishda tilning madaniy, ijtimoiy va psixologik aspektlarini hisobga olish zarur. Bu, tilshunoslik va madaniyatshunoslik yondashuvlaridan foydalanib amalga oshirilishi mumkin.

Milliy kontent yaratishda adabiy til me'yorlariga rioya qilish tilning estetik va madaniy funksiyalarini samarali amalga oshirishga xizmat qiladi. Tilshunoslik va madaniyatshunoslikning interdisipliner yondashuvlari orqali milliy identifikatsiyaning ifodalovchi, tilni to'g'ri va madaniy jihatdan samarali ishlatish imkoniyatlari yaratiladi. Milliy kontentda til me'yorlariga rioya qilish ijtimoiy va madaniy qadriyatlarni to'g'ri ifodalashga, shuningdek, tilni doimiy ravishda rivojlantirishga yordam beradi. Bu esa milliy madaniyatni saqlash va mustahkamlashda muhim rol o'ynaydi.

Milliy kontentda til me'yorlariga rioya qilish, tilning sifatli ishlatilishini ta'minlabgina qolmay, balki ijtimoiy va madaniy qadriyatlarning to'g'ri ifodalashini ham muhim darajada oshiradi. Har bir millat o'z tilida nafaqat o'zining tarixiy, madaniy, va ijtimoiy xususiyatlarini aks ettiradi, balki bu til orqali o'zining ruhiy holati, qadriyatlar va an'analari haqida ham ma'lumot beradi. Til me'yorlariga rioya qilishning bu jarayondagi ahamiyati ikki tomonlama: bir tomondan, u tilni rivojlantirishga yordam beradi, ikkinchi tomondan esa madaniy qadriyatlarni til orqali to'g'ri ifodalashga imkon yaratadi.

Ijtimoiy qadriyatlarni to'g'ri ifodalash: Milliy kontentda adabiy til me'yorlariga rioya qilish, ijtimoiy qadriyatlarni aks ettirishning muhim vositasi bo'lib xizmat qiladi. Til o'zini ifodalayotgan ijtimoiy guruhning urf-odatlarini, an'analari va qadriyatlarini tasvirlaydi. Masalan, o'zbek xalqining "mehmondo'stlik" yoki "salomlashish" madaniyati tilga to'g'ri aks ettirilsa, bu faqatgina tilni emas, balki o'sha xalqning ijtimoiy hayotining o'ziga xos tomonlarini ham

anglatadi. Agar bu tushunchalar noto'g'ri ishlatilsa, masalan, "mehmondo'st" so'zi salbiy kontekstda ishlatilsa, u o'sha xalqning ijtimoiy munosabatlarini noto'g'ri aks ettirishga olib keladi. Bunday holatlarda, til nafaqat madaniy qadriyatlarni buzadi, balki u ijtimoiy identifikatsiyaning noto'g'ri talqin etilishiga olib kelishi mumkin.

Foydalanilgan adabiyotlar:

1. Shumilov, S. (2017). Til va madaniyatning o'zaro bog'liqligi: Milliy identifikatsiya va kontent yaratish (p. 123). Toshkent: O'zbekistan Yozuvchilar Uyi.
2. Avtandilov, M. (2018). Adabiy til me'yorlari va madaniyat (p. 45). Moskva: Akademiya.
3. Saussure, F. de (1916). Course in General Linguistics. McGraw-Hill.
4. Halliday, M. A. K. (1978). Language as Social Semiotic: The Social Interpretation of Language and Meaning. Edward Arnold.

MULTIMEDIA KORPUSLARINING TIL O'QITISHDAGI O'RNI VA AHAMIYATI

Abduraxmonova Nilufar Zaynobiddin qiziO'zbekiston Milliy universiteti professori
Filologiya fanlari doktoriE-mail: n.abduraxmonova@nuu.uz**Rajabova Karomat Chorshanbi qizi**O'zbekiston Milliy universiteti
Kompyuter lingvistikasi mutaxassisligi magistri
E-mail: Rajabovakaromat9@gmail.com

Annotatsiya: Ushbu maqolada multimedia korpuslari, ularning tuzilishi, jahonda yaratilgan yirik multimedia korpuslarining imkoniyatlari va zamonaviy til o'qitish jarayonidagi o'rni keng yoritilgan. Multimedia korpuslarining til o'rganishdagi imkoniyatlari, o'quvchilarning multimodal til kompetensiyalarini rivojlantirishdagi roli ilmiy manbalar asosida tahlil qilingan.

Kalit so'zlar: Media, videomatn, kontent, ta'limiy korpus, interfeys, subkorpus.

Annotation: This article extensively covers multimedia corpora, their structure, capabilities of large multimedia corpora created in the world, and their role in the modern language teaching process. The possibilities of multimedia corpora in language learning, the role of students in the development of multimodal language competencies are analyzed based on scientific sources.

Keywords: Media, video text, content, educational corpus, interface, subcorpus.

Аннотация: В данной статье широко освещены мультимедийные корпуса, их структура, возможности крупных мультимедийных корпусов, созданных в мире, и их роль в современном процессе обучения языку. На основе научных источников проанализированы возможности мультимедийных корпусов в изучении языка, роль учащихся в развитии мультимодальных языковых компетенций.

Ключевые слова: Медиа, видеотекст, контент, образовательный корпус, интерфейс, субкорпус.

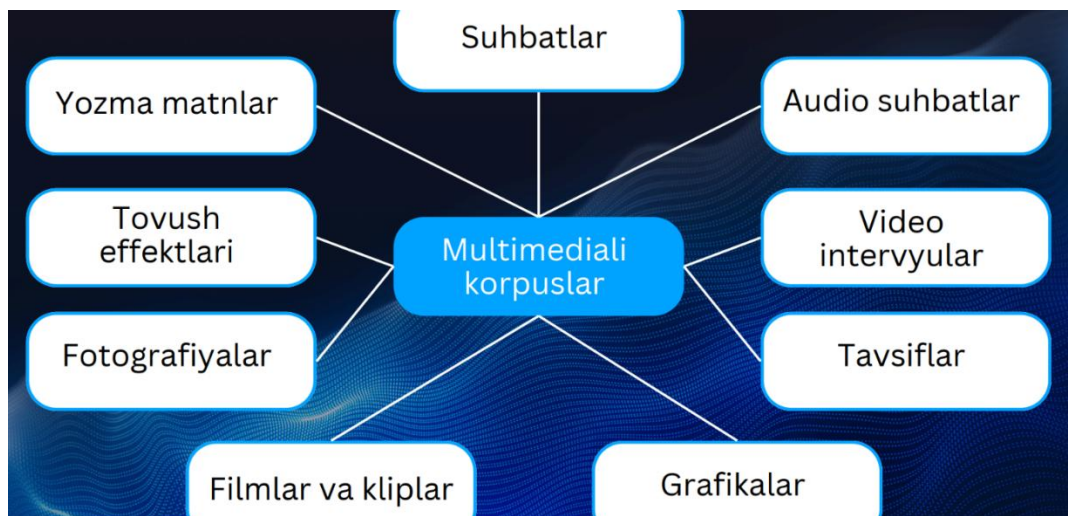
Kirish. Axborot texnologiyalari va sun'iy intellektning jadal rivojlanishi tilshunoslik va ta'lim tizimida yangi yondashuvlarni talab etmoqda. Xususan, multimedia korpuslari tabiiy tilning multimodal xususiyatlarini tahlil qilish va ularni ta'lim jarayonida samarali qo'llash uchun keng imkoniyat yaratmoqda. Multimedia korpuslar yordamida o'quvchilar nafaqat matnli, balki og'zaki, audio, video va kontekstual axborotlar asosida ham til kompetensiyalarini rivojlantirishi mumkin.

Multimedia korpusi — bu matn, audio, video, tasvirlar va boshqa turdagi medialarni tizimli va anotatsiyalangan holatda jamlagan ma'lumotlar to'plamidir.

Jahon korpus lingvistikasida multimediali korpuslar ko'plab yaratilgan. Xususan, ingliz tilining zamonaviy shaklini o'rganishda qo'llaniladigan – COCA korpusi [<https://www.english-corpora.org/coca/>], English-Corporadagi film va seriallar korpusi shular jumlasidandir. Filmlar korpusi Mark Devis tomonidan yaratilgan [<https://www.english-corpora.org/movies/x1.asp>]. Fransuz tilining ESLO korpusi [<http://eslo.huma-num.fr/>] asosan, og'zaki nutqni o'rganishga asoslangan korus hisoblanadi. Multimedia Rus Korpusi (MURKO)[<https://ruscorpora.ru/>] rus tili milliy korpusining subkorpusi bo'lib, unda audio va video qismlarga moslashtirilgan og'zaki ruscha matnlar mavjud. Turkologiyada korpusshunoslik va multimedia korpuslari bo'yicha Turk Milliy Korpusi (TUD)[<https://www.tnc.org.tr/>], Xakas tili elektron korpus[<http://khakas.altai.ru/>], Qozoq tili zamonaviy so'zlashuv uslubining multimedia korpusi

dasturi [https://research.nu.edu.kz/en/projects/multimedia-corpus-of-modern-spoken-kazakh-language] va boshqalarni misol keltirishimiz mumkin.

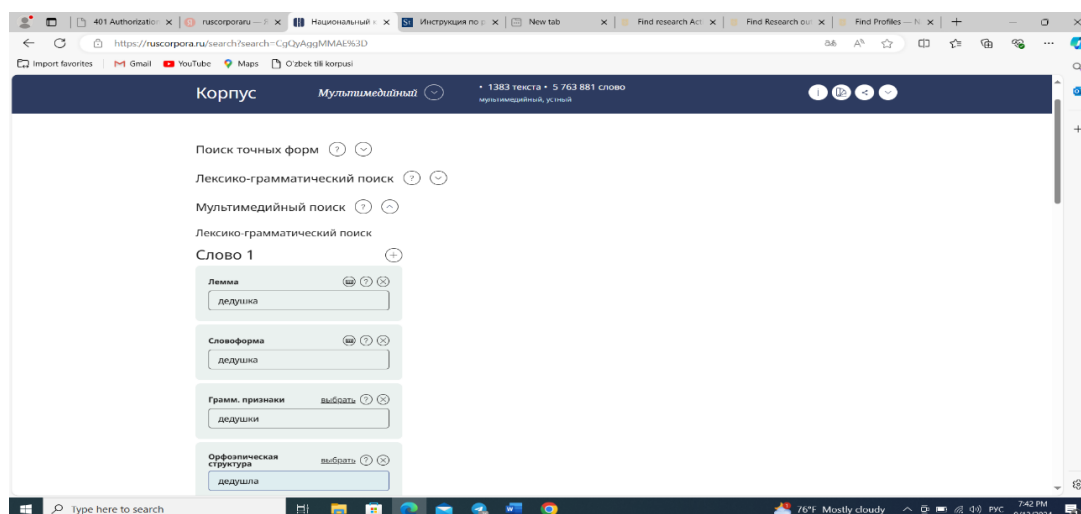
Multimediali korpuslar quyidagi asosiy komponentlarni o'z ichiga oladi:



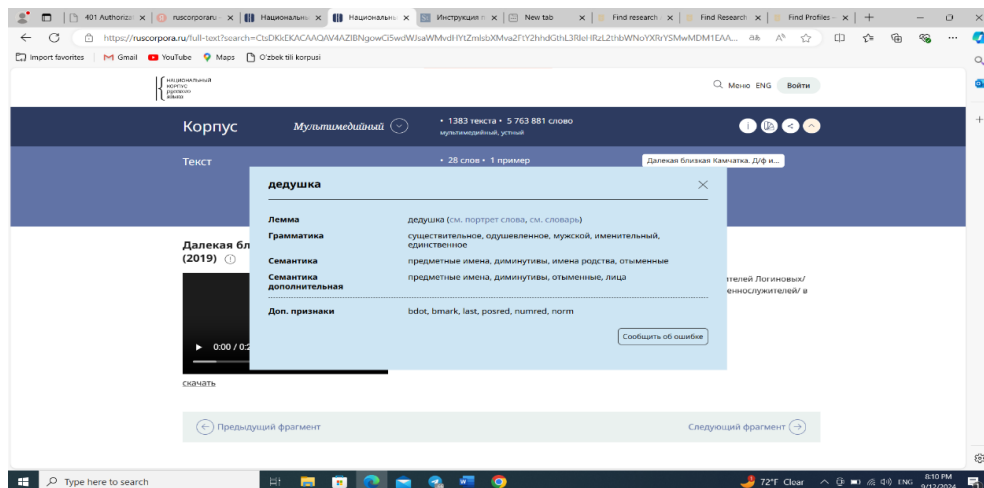
Multimedia Rus Korpusi (MURKO)[<https://ruscorpora.ru/>] rus tili milliy korpusining subkorpusi hisoblanib, unda tegishli audio va video qismlarga moslashtirilgan og'zaki ruscha matnlar mavjud. MURKO og'zaki nutqni nafaqat uning yozma fiksatsiyasidan, balki haqiqiy tovush va imo-ishora hamrohligidan foydalangan holda o'rganishga imkon beradi. MURKO birinchi navbatda fonetika, imo-ishoralar bo'yicha mutaxassislar uchun mo'ljallangan, lekin u rus tili o'qituvchilari uchun, birinchi navbatda, undan tashqari rus tilini chet tili sifatida o'rganuvchilarga ham foydali bo'ladi. Multimedia korpusi rus tili milliy korpusining ajralmas qismi sifatida tilni o'rganish va tahlil qilishda samarali resurs hisoblanadi.

https://studiorum.ruscorpora.ru/murco_instruction/ – bu korpus bilan ishlash bo'yicha alohida yo'riqnoma portali bo'lib, 2017-yili ishlab chiqilgan.

Multimedia rus korpusidagi asosiy qidiruv sahifasi rus tili milliy korpusining boshqa modullari kabi tuzilgan, ammo kerakli qidiruv imkoniyatlarini beradigan qo'shimcha parametrlarga ega. Misol uchun MURKO[https://studiorum.ruscorpora.ru/murco_instruction/] da leksik-grammatik va semantik qidiruv parametrlari mavjud:



Multimedia korpusidagi so'zlar grammatik, leksik, semantik jihatdan tahlilga tortilgan. Lemma bo'yicha qidirsak, uning lug'atdagi shakli, grammatik jihatdan qaysi so'z turkumiga oidligi, semantik tomonlari ko'rsatiladi:

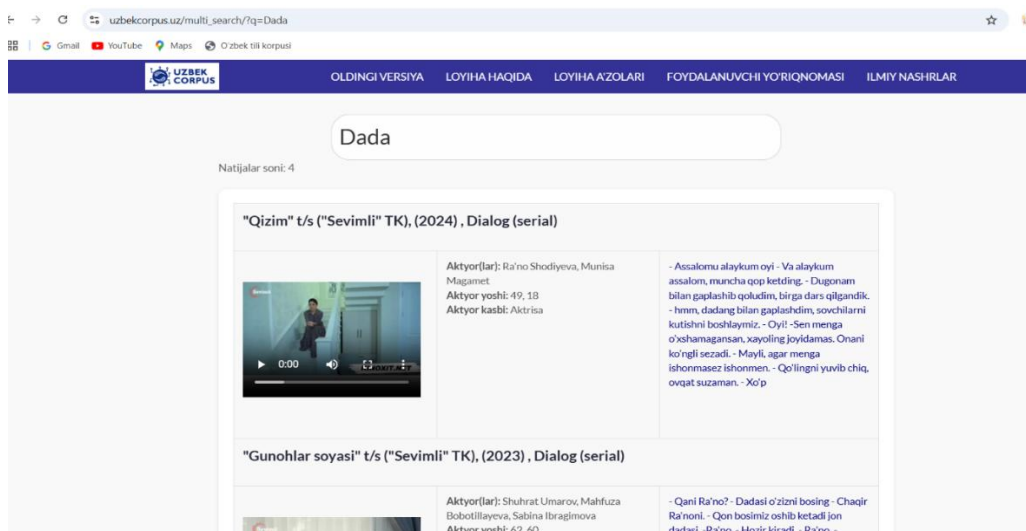


Korpusning ushbu jihatlari til o'rganuvchilarga tilning fonetika, leksika, sintaksis, va semantikasini o'rganish imkoniyatini kengaytiradi.

O'zbekistonda ham multimedia korpuslarini tuzish va uni ta'limga yo'naltirish borasida salmoqli loyihalar amalga oshirilmoqda. Filologiya fanlari doktori, professor Nilufar Abduraxmonova tomonidan yaratilgan uzbekcorpus.uz [uzbekcorpus.uz] saytida uning subkorpusi sifatida multimedia korpusi yaratildi.

Korpus interfeysi [<https://uzbekcorpus.uz/multimedia>] bu uzbekcorpus.uz saytidagi subkorpus sifatida ta'limiy korpus hisoblanadi. Ushbu ta'limiy korpus o'zbek tilini ikkinchi va xorijiy til sifatida o'rganuvchilarga til haqidagi grammatik ma'lumotlarni berish imkoniyatini yaratadi. Shu boisdan korpus interfeysi barcha foydalanuvchilar uchun qulay va sodda ko'rinishda yaratilgan bo'lib, bu qidiruv jarayonini tezlashtirish imkonini beradi.

Korpus qidiruv interfeysi PYTHON dasturlash tilida amalga oshirilgan. Python Artificial Intelligence (Sun'iy intellekt) va Data Science (Ulkan ma'lumotlar bilan ishlash) sohalarining tili hisoblanadi. Interfeysda qidirilgan so'zning videomatni o'ng tomonda, videokontent esa chap tomonda joylashtirilgan. Korpus interfeysida videokontentlarning metama'lumoti ham joylashtirilgan. Quyidagi rasmda videokontentlarning metama'lumoti, sarlavhasi, janri va paragraflarini ko'rishimiz mumkin:



Ushbu yaratilgan multimedia korpusida morfologik teglash tizimi mavjud bo'lib, til o'rganuvchisiga fonetik va grammatik ma'lumotlarni berish imkoniyatini yaratadi:

→ uzbekcorpus.uz/multi_search?q=Dars# ☆

Gmail YouTube Maps O'zbek tili korpusi

UZBEK CORPUS OLDINGI VERSIYA LOYIHA HAQIDA LOYIHA A'ZOLARI FOYDALANUVCHI YO'RIQNOMASI ILMIY NASHRLAR

Dars

Natijalar soni: 1

"Qizim" t/s ("Sevimli" TK), (2024), Dialog (serial)



Aktor(lar): Ra'no Shodiyeva, Munisa Magamet
Aktor yoshi: 49, 18
Aktor kasbi: Aktrisa

- Assalomu alaykum oyi - Va alaykum assalom, muncha qop ketding. - Dug bilan gaplashib qoludim, birga dars. - hmn, dadang bilan gaplashdim, sov kutishni boshlaymiz. - Oyi! - Sen men o'xshamagansan, xayoling joyidamas ko'ngli sezadi. - Mayli, agar menga ishonmasez ishonmen. - Qo'lingni yu ovqat suzaman. - Xo'p

dars	
Lemma	dars lug'atda ko'rish
So'z turkumi	NOUN
Morfologik teg	Noun+P3+SG

Xulosa. Multimedia korpuslari lingvistika va til o'qitishda ilg'or texnologik vosita bo'lib, tabiiy til muhitini yaratish, multimodal kompetensiyalarni rivojlantirish va madaniyatlararo kommunikatsiya ko'nikmalarini shakllantirish imkonini beradi. Multimedia korpuslari asosida tashkil etilgan o'quv mashg'ulotlari o'quvchilarning til va kommunikativ kompetensiyalarini oshirishda yuqori samaradorlikka ega. Shu bois, multimedia korpuslari asosida yaratilgan o'quv materiallari va ta'lim usullarini keng joriy etish maqsadga muvofiq.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. O'zbek tili elektron korpusining kompyuter modellari. Monografiya. Toshkent-2021. – 400 b.
2. Abduraxmonova N. Korpus lingvistikasi. Darslik. T., 2023. – 346 b
3. Abduraxmonova N. Kompyuter lingvistikasi (darslik), Toshkent: Nodirabegim, 2021. – 397 b.
4. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE
5. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240)
6. Mengliyev B., Hamrayeva Sh. Korpus lingvistikasi: korpus tuzish va undan foydalanish. – Toshkent: Globeedit. 2020. – 50 b.
7. Ide, N., & Macleod, C. (2001). The American National Corpus: A standardized resource of American English. Proceedings of Corpus Linguistics 2001.
8. Johnston, M. (2009). Multimodal Human-Computer Interaction from Signals to Interaction. Cambridge University Press
9. Baldassarri, S., et al. (2009). A Multimodal Corpus for Emotion Analysis in Spanish. Proceedings of LREC 2008.
10. Cao, H., et al. (2014). CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset. IEEE Transactions on Affective Computing.
11. Wittenburg, P., et al. (2006). ELAN: a professional framework for multimodality research. Proceedings of LREC 2006.
12. <http://uzbekcorpus.uz>
13. <https://uzbekcorpus.uz/multimedia>
14. <https://www.english-corpora.org/coca/>

15. <http://eslo.huma-num.fr/>
16. <https://ruscorpora.ru/>
17. <https://research.nu.edu.kz/en/projects/multimedia-corpus-of-modern-spoken-kazakh-language>
18. https://studiorum.ruscorpora.ru/murco_instruction/

CHET ELLIK TALABALARNI O'QITISHDA KORPUSGA ASOSLANGAN YONDASHUV

Shamsiyeva Rayhona Saidahmad qizi

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti
Jurnalistika va o'zbek filologiyasi fakulteti 2-bosqich talabasi

E-mail: rayhonashamsiyeva2@gmail.com

Annotasiya: Ushbu maqolada chet tilini o'rganuvchilarni samarali o'qitishda korpusga asoslangan yondashuvning o'rni va afzalliklari tahlil qilinadi. Mavjud ilmiy adabiyotlar asosida bu metodikaning til kompetensiyalarini (leksik, grammatik, yozma va og'zaki nutq) rivojlantirishdagi samaradorligini ko'rsatib beriladi. Shuningdek, maqolada korpus lingvistikasi vositalari va ularning dars jarayonidagi amaliy tatbiqlari tahlil qilinadi. Tadqiqot natijalari korpusga asoslangan yondashuvni o'quv metodikasiga integratsiyalashning dolzarbligini asoslaydi.

Kalit so'zlar: korpus lingvistikasi, chet tili o'qitish, til kompetensiyasi, o'quv materiallari, metodika.

Abstract: This article analyzes the role and advantages of the corpus-based approach in the effective teaching of foreign language learners. Based on the existing scientific literature, the author demonstrates the effectiveness of this methodology in developing language competencies (lexical, grammatical, written and oral). The article also analyzes the tools of corpus linguistics and their practical applications in the teaching process. The results of the study justify the relevance of integrating the corpus-based approach into teaching methods.

Keywords: corpus linguistics, foreign language teaching, language competence, educational materials, methodology.

Аннотация: В статье анализируются роль и преимущества корпусного подхода в эффективном обучении изучающих иностранные языки. Автор на основе имеющейся научной литературы демонстрирует эффективность данной методики в развитии языковых компетенций (лексических, грамматических, письменных и устных). В статье также анализируется инструментарий корпусной лингвистики и его практическое применение в учебном процессе. Результаты исследования подтверждают актуальность внедрения корпусного подхода в методики обучения.

Ключевые слова: корпусная лингвистика, преподавание иностранных языков, языковая компетенция, учебные материалы, методика.

So'nggi yillarda chet tillarini o'qitishda innovatsion yondashuvlar keng rivojlanmoqda. Shulardan biri — korpusga asoslangan yondashuv bo'lib, u til o'rgatish jarayonida real til materiali asosida o'qitishni taklif etadi. An'anaviy metodlarda ko'proq til strukturasi va nazariy qoidalar asosiy o'rinda bo'lsa, korpus lingvistikasi vositalari orqali tilni qanday ishlatilishini amaliy misollar orqali o'rgatish imkoniyati mavjud bo'ladi. Bu ayniqsa, chet ellik o'rganuvchilar uchun tilni tabiiy muhitda o'zlashtirishda muhim rol o'ynaydi. Ushbu maqolada korpus lingvistikasi yondashuvining nazariy asoslari, tarixiy shakllanishi, zamonaviy korpus turlari va ularning o'qitishdagi amaliy tatbiqlari yoritiladi.

Korpus lingvistikasi — bu real hayotda ishlatilgan matnlar asosida til birliklarini tahlil qiluvchi ilmiy yo'nalish bo'lib, u empirik ma'lumotlarga asoslanadi. Korpus — bu muayyan tilga oid raqamlashtirilgan va maxsus tahliliy dasturlar yordamida o'rganiladigan matnlar to'plamidir. Ushbu korpuslar morfologik, sintaktik va semantik belgilar bilan belgilanishi mumkin. Bu borada Jon Sinklarning xizmatlari alohida ahamiyatga ega. Bu olim korpus lingvistikasining asoschilaridan biri bo'lib, uning 1991-yilda chop etilgan "Corpus, Concordance, Collection" nomli

asari ushbu sohaning rivojlanishida muhim rol o'ynagan. Asarda Sinkler so'zlarning ma'nosi ularning kontekstda qanday ishlatilishiga bog'liq ekanligini ta'kidlaydi. Uning fikricha, so'zlar o'zaro bog'liq holda, ya'ni kollakatsiyalar orqali ma'no hosil qiladi. Bu yondashuv til o'rganishda so'zlarni izolyatsiyada yoki ma'lum bir qat'iy qoidalar bilan emas balki ularning tabiiy kontekstida o'rganish mumkinligini ko'rsatadi. Duglas Biberda bu biroz boshqacha uslubda. Korpus lingvistikasi va funksional tilshunoslik sohalarida taniqli bo'lgan bu olimning 1998-yilda chop etilgan "Corpus Linguistics: Investigating Language Structure and Use" kitobida grammatika va stilistikani tahlil qilishda statistik yondashuvlarni ilgari suradi. Biber turli matn janrlarini tahlil qilib ularning lingvistik xususiyatlarini statistik usullar bilan o'rganadi. Uning yondashuvi matnlarning funksional xususiyatlarini aniqlashda korpus lingvistikasining imkoniyatlarini namoyish etadi. Bundan tashqari Toni Makenri va Endryu Hardining izlanishlari ham muhim ahamiyat kasb etadi. Ularning 2011-yilda nashr etilgan "Corpus Linguistics: Method, Theory and Practice" asarida korpus tahlilining nazariy va amaliy uslublarini chuqur ochib beradi. Bunda korpus lingvistikasining metodologik asoslarini, turli usullarini va ularning amaliy qo'llanilishini batafsil yoritiladi. Bundan anglashiladiki korpus lingvistikasi orqali tilshunoslar nafaqat til birliklarini tahlil qiladilar, balki til o'zgarishi, nutq odatlari va kontekstual ma'nolarni ham o'rganadilar. Bu esa chet tillarini o'rganuvchi talabalar uchun tabiiy til modelini taqdim eta oladilar. Shu bilan birga bu turdagi korpuslar yaratishda o'rganuvchilarning ta'lim muhiti, yoshi, ona tilisi, ta'lim bosqichi kabi belgilari hisobga olinadi.⁹

Chet tillarini o'qitishda korpuslardan foydalanish 1990-yillardan boshlab jadal rivojlandi. Dastlab, universitetlar va tadqiqot markazlarida ishlab chiqilgan korpuslar lingvistik tahlil uchun xizmat qilgan bo'lsa, keyinchalik ularning o'quv jarayoniga integratsiyasi boshlandi. Masalan, British National Corpus (BNC) — ingliz tilining zamonaviy variantini ifodalovchi 100 milliondan ortiq so'zdan iborat korpusidir. Korpus tarkibiga turli janrdagi yozma matnlar (ilmiy, badiiy, rasmiy, ommaviy axborot vositalari matnlari) hamda og'zaki nutq namunalari kiritilgan. Talabalar bu korpus yordamida keng qo'llanuvchi iboralarni, grammatik konstruksiyalarni va kontekstual misollarni tahlil qilish imkoniga ega bo'ladilar. Ayniqsa, kollakatsion o'zgarishlarda BNC yirik manba sifatida xizmat qiladi. Misol uchun ingliz tilida "make a decision" (qaror qabul qilmoq) iborasi bor. Bu ibora BNCda minglab marta uchraydi, ammo grammatik jihatdan mumkin bo'lsa-da, "do a decision" iborasi deyarli ishlatilmaydi. Ko'rinadiki, ikkinchi shakl deyarli ishlatilmaydi. Bu esa til o'rganuvchilar uchun real qo'llaniladigan til va sun'iy holatlar o'rtasidagi farqni anglashda muhim vositadir. Shuningdek, COCA (Corpus of Contemporary American English) va Cambridge Learner Corpus kabi resurslar xorijiy talabalar uchun o'zbek, rus, ingliz va boshqa tillardagi o'quv resurslarini yaratishda asos bo'lmoqda. Bu korpus har yili yangilanib boradi va televizion matnlar, gazetalar, bloglar, ilmiy maqolalar va og'zaki nutqni qamrab oladi. COCA'ning muhim jihati – unda janrlararo, vaqt bo'yicha va uslubiy farqlarni solishtirish imkoniyati mavjudligidir. Shuningdek, COCA'ning veb-interfeysi orqali foydalanuvchi so'zning konkordans, ya'ni kontekstdagi ishlatilishini, u bilan eng ko'p birga keladigan so'zlarni (collocates), grammatik strukturadagi roli va vaqt o'tishi bilan qo'llanilish chastotasini tahlil qilishi mumkin. Misol uchun, "awesome" (ajoyib) so'zining qo'llanilishi COCA korpusida 1990-yillardan keyin keskin oshganini ko'rish mumkin. Bu orqali biz slang va uslubiy o'zgarishlarni aniqlashimiz mumkin.

Tarixan korpuslar chet tilini o'rgatuvchi metodistlar uchun yangi yondashuvlarni shakllantirishga xizmat qilgan. O'qituvchilar endi real tildan foydalanishiga tayanib, o'quv materiallarini yaratishmoqda. Shu bilan birga korpusga asoslangan yondashuvda nazariy bilimlar bilan bir qatorda, korpus materiallarini tahlil qilishga imkon beruvchi dasturiy vositalar ham muhim ahamiyat kasb etadi. Bunday dasturlar yordamida tilshunoslar va o'qituvchilar real til faktlarini tizimli tahlil qilishi, chet ellik til o'rganuvchilarga til birliklarining kontekstdagi

⁹ Abduraxmonova N. Korpus lingvistikasi Toshkent-2023. -357b.

ishlatilishiga asoslangan mashg'ulotlarni tashkil etishlari mumkin. Jumladan, AntConc, Sketch Engine va WordSmith Tools kabi dasturlar jarayoni sezilarli darajada yengillashtiradi va boyitadi.

Dastlab AntConc dasturiga to'xtaladigan bo'lsak, bu dastur bepul, ochiq manbali bo'lib, u orqali o'qituvchi yoki o'rganuvchi ma'lum bir so'z yoki iboraning korpusdagi barcha ishlatilish hollarini tezkorlik bilan ko'rish imkoniga ega bo'ladi. Konkordans funksiyasi orqali til birliklarining oldi va orqa kontekstidagi so'zlarni ko'rish mumkin. "Collocates" funksiyasi orqali esa so'zning eng ko'p birga keladigan juftliklarini ajratib beradi. Masalan, O'zbek Milliy Korpusi materiallari asosida "kitob" so'zining chap va o'ng tarafidan 5 so'z oralig'idagi kollokatsiyalar AntConc orqali tahlil qilinsa, quyidagi natijalar kuzatiladi:

So'z juftligi	Kollokatsiya chastotasi
kitob o'qimoq	124
kitob do'koni	47
kitob yozmoq	35
kitob sovg'a qilmoq	21
kitobxon	89

Bu natijalar shuni ko'rsatadiki, o'zbek tilida "kitob" so'zi eng ko'p "o'qimoq", "do'kon", "yozmoq" kabi so'zlar bilan bog'liq. Shu asosda o'qituvchi o'quvchilarga semantik jihatdan gap tuzish mashqlarini berishi, yoki o'z matnlarini korpusga tayanib tahlil qildirish imkoniyatiga ega bo'ladi.

Yana bir qulay platforma – bu Sketch Engine bo'lib, u yanada murakkab lingvistik tahlillarni amalga oshirish imkonini beradi. Ayniqsa, "Word Sketch" deb nomlanuvchi funksiya orqali foydalanuvchi har qanday so'zning sintaktik roli, u bilan keladigan fe'l, sifat, otlar tizimini grafik shaklda ko'rish mumkin. Bu esa murakkab grammatik strukturalarni yodlash emas, balki tushunib o'zlashtirish jarayonini kuchaytiradi. Shu bilan birga, ushbu platforma foydalanuvchiga o'zining korpusini yaratish va tahlil qilish imkonini ham beradi. Shuningdek, WordSmith Tools ham muhim dasturlardan biri bo'lib ko'proq akademik tahlil va yozma matnlar ustida ishlashda foydalidir. U foydalanuvchiga matnlar bo'yicha so'zlar ro'yxatini (wordlist) yaratish, chastotani tahlil qilish, hamda boshqa matnlar bilan solishtirib kalit so'zlar(keywords)ni aniqlash imkonini beradi. Masalan, ilmiy maqolalarda eng ko'p ishlatiladigan frazalar va iboralarini aniqlash orqali o'quvchilar akademik yozuvga xos uslubni o'zlashtira oladilar. Bundan tashqari, talabalarning yozgan matnlarini WordSmith orqali korpus bilan taqqoslash ularning grammatik xatolarini imkoni mavjud.

Yuqorida keltirilgan dasturlar orqali korpusga asoslangan o'rgatish nafaqat statistik tahlilga asoslangan, balki kontekstda o'rganishga tayangan yondashuv sifatida taqdim etiladi. Bu esa tilni sun'iy tarzda emas, balki hayotdagi holatlarda qanday ishlatilishini chuqur anglashda muhim ahamiyat kasb etadi. Ayniqsa, chet ellik o'rganuvchilar uchun bu yondashuv til o'zlashtirish jarayonidagi vaqtni tejashda foydali bo'ladi. Buning natijasi o'laroq o'tkazilgan ilmiy tadqiqotlar korpusga asoslangan yondashuvning samaradorligini ko'rsatmoqda. Masalan, O'Keefe, McCarthy va Carter (2007) o'tkazgan tajribada korpus bilan ishlagan talabalar grammatikani an'anaviy darsga qaraganda 30% tezroq va puxtaroq o'zlashtirgani

aniqlangan.¹⁰Yana bir muhim tadqiqot sifatida Ludeling va Kyto (2008) tahririda chop etilgan “Corpus Linguistics: An International Handbook” asaridagi korpus asosida tilshunoslik tadqiqotlari keltiriladi. Ularda chet ellik o’rganuvchilar tomonidan yozilgan matnlar tahlili orqali, o’zlashtirilgan va o’zlashtirilmagan grammatik shakllar aniqlangan. Masalan, ingliz tilini ikkinchi til sifatida o’rganuvchilarning yozma nutqida present perfect zamonidan kam foydalanish tendensiyasi kuzatilgan. Bu holat korpusdagi avtomatik analiz orqali aniqlandi va o’quv dasturlariga tegishli grammatik modullarni kuchaytirish zarurligini ko’rsatdi.

Korpusga asoslangan yondashuv chet tilini samarali va chuqur o’zlashtirish uchun zamonaviy, ilmiy asoslangan va amaliy jihatdan qulay metod hisoblanadi. Bu yondashuv o’rganuvchini haqiqiy kontekstda tilni o’rgatib, tahliliy fikrlashga undaydi. O’zbekistonda ham bu yo’nalishdagi izlanishlar endi boshlanayotgan bo’lsa-da, mavjud imkoniyatlar asosida rivojlanib bormoqda.

Xulosa qilib aytganda chet ellik talabalarni samarali til o’rganishga yo’naltirishda korpusga asoslangan yondashuv zamonaviy va ilmiy asoslangan metod sifatida katta ahamiyatga ega. Bu yondashuv tilni haqiqiy kontekstda o’rgatish, grammatik va leksik bilimlarni chuqur o’zlashtirish, yozma va og’zaki nutqni takomillashtirish imkonini beradi. Korpus lingvistikasi vositalari, xususan, WordSmith Tools, Sketch Engine, AntConc kabi dasturlar orqali o’quvchilar til birliklarining real holatdagi qo’llanilishi, ularning kontekstual ma’nolari, grammatik xatolari va uslubiy xususiyatlarini chuqur tahlil qilish imkoniga ega bo’ladilar. Tadqiqotlar shuni ko’rsatadiki, korpus asosida til o’rgatish nafaqat statistik ma’lumotlarga asoslanadi, balki tahliliy fikrlash, kontekstda tushunish va til kompetensiyalarini shakllantirishda muhim rol o’ynaydi. Shu sababli korpusga asoslangan yondashuv bugungi kunda chet tillarini o’rgatishda istiqbolli metodlardan biri sifatida qaraladi.

Foydalanilgan adabiyotlar:

1. Sinclair J. Corpus, Concordance, Collocation. — Oxford: Oxford University Press, 1991. — 200 p.
2. Biber D. Corpus Linguistics: Investigating Language Structure and Use. — Cambridge: Cambridge University Press, 1998. — 320 p.
3. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. — Cambridge: Cambridge University Press, 2011. — 290 p.
4. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE
5. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
6. Abduraxmonova N. Korpus lingvistikasi Toshkent-2023. -357b.
7. O’Keeffe A., McCarthy M., Carter R. From Corpus to Classroom: Language Use and Language Teaching. — Cambridge: Cambridge University Press, 2007. — 282 p.
8. Kyuto T. Korpusga asoslangan o’rganish: til o’rgatuvchilar uchun metodologik imkoniyatlar // Computer Assisted Language Learning. – 2020. – T. 33, № 5–6. – B. 558–582. – <https://10.1080/09588221.2020.173469>

¹⁰ O’Keeffe A., McCarthy M., Carter R. From Corpus to Classroom: Language Use and Language Teaching. — Cambridge: Cambridge University Press, 2007. — 282 b.

SENTIMENT TAHLILDA TASNIFLASH SXEMASI

Abdullayev Abdulla Quranbayevich

"Urganch innovatsion university" NTM ilmiy tadqiqotlar,
innovatsiyalar va ilmiy pedagogik kadrlarni tayyorlash bo'limi boshlig'i.

E-mail: abdulla_abdullayev9270@mail.ru

ORCID: 0009-0007-9037-1364

Annotatsiya: O'zbek tilida sentiment tahlilni amalga oshirishda tasniflash sxemasini ishlab chiqish tilning morfologik murakkabligi, kodli almashish va ma'lumotlar tanqisligi kabi muammolarni hisobga olishni talab qiladi. Ushbu maqolada, sentiment tahlilda tasniflash sxemasini ishlab chiqish uchun zarur bo'lgan asosiy bosqichlar – ma'lumotlarni yig'ish, tozalash va tayyorlash, xususiyatlar ajratish, modelni tanlash va o'qitish, baholash va optimallashtirish, shuningdek, modeldan foydalanish – ilmiy asosda tahlil qilinadi.

Kalit so'zlar: Sentiment tahlil, dataset yaratish, ma'lumotlarni yig'ish, preprocessing, xususiyatlarni ajratish, modelni tanlash, modelni o'qitish, baholash, optimallashtirish, joriy etish.

Аннотация: Разработка схемы классификации при проведении сентиментального анализа в узбекском языке требует учета таких проблем, как морфологическая сложность языка, обмен кодами и дефицит информации. В данной статье на научной основе анализируются основные этапы, необходимые для разработки схемы классификации в сентиментальном анализе - сбор, очистка и подготовка данных, выделение характеристик, выбор и обучение модели, оценка и оптимизация, а также использование модели.

Ключевые слова: Сентиментальный анализ, создание набора данных, сбор данных, препроцессинг, выделение свойств, выбор модели, обучение модели, оценка, оптимизация, внедрение.

Abstract: When conducting sentiment analysis in the Uzbek language, the development of a classification scheme requires taking into account such problems as the morphological complexity of the language, code exchange, and data scarcity. In this article, the main stages necessary for developing a classification scheme in sentiment analysis are analyzed on a scientific basis - data collection, purification and preparation, characterization, model selection and training, evaluation and optimization, as well as the use of the model.

Keywords: Sentiment analysis, dataset creation, data collection, preprocessing, property separation, model selection, model training, evaluation, optimization, implementation.

Kirish. Zamonaviy sun'iy intellekt texnologiyalarining rivoji bilan bir qatorda, matnli ma'lumotlarni avtomatik tahlil qilishning ahamiyati ortib bormoqda. Sentiment tahlil (his-tuyg'u tahlili) ushbu sohaning asosiy yo'nalishlaridan biri bo'lib, u mijozlar sharhlaridan tortib siyosiy nutqlarigacha turli kontekstlarda foydalanuvchi nuqtai nazarini aniqlashga qaratilgan. O'zbek tilida sentiment tahlilni amalga oshirishda tilning morfologik murakkabligi, kodli almashish (rus/ingliz tillaridan so'zlar qo'shilishi) va yetarli hajmdagi annotatsiyalangan ma'lumotlar to'plamining yo'qligi kabi muammolar mavjud. Xalqaro tadqiqotlar shuni ko'rsatadiki, tilning strukturaviy xususiyatlarini hisobga olgan holda ishlab chiqilgan tasniflash sxemasi model samaradorligini keskin oshirishi mumkin[Tang, 2015, 12; Cambria, 2020,53]. O'zbek tiliga xos yechimlarni shakllantirish uchun maqolada 6 bosqichli metodologiya taklif etiladi: **ma'lumotlarni yig'ish, tozalash, xususiyatlar ajratish, modelni o'qitish, baholash va amaliy joriy etish.** O'zbek tiliga mo'ljallangan ushbu sxema nafaqat lingvistik komplekslikni bartaraf etish, balki xorijiy tillar uchun ishlab chiqilgan yechimlarni mahalliy kontekstga moslashtirishga imkon beradi. Tadqiqotning amaliy ahamiyati shundaki, u o'zbek NLP hamjamiyatiga sentiment tahlil vositalarini ishlab chiqish uchun metodologik asos yaratadi. Kelajakda ushbu sxemani ko'p tilli

transformer modellar (masalan, LaBSE) bilan integratsiyalash orqali universal yechimlarni ishlab chiqish mumkin.

Ushbu tadqiqotda sentiment tahlilning tasniflash sxemasi nazariy va amaliy jihatdan tahlil qilinib, uning yaratilishi tilning murakkab leksik, semantik va kontekstual xususiyatlarini inobatga olishni talab qiladi. Ilmiy izlanishlarda[Pang va Lee, 2008,27] leksikon asosidagi metodlarning tezkorligi va soddaligini ta'kidlagan. Shuningdek, mashinani o'qitish va chuqur o'rganish metodlari yordamida murakkab semantik munosabatlarni aniqlash imkoniyatlarini namoyish etish mumkinligi, bu esa sentiment tahlilda yuqori aniqlikka erishishda muhim omil sifatida ko'rilishiga urg'u berilgan[Socher, 2013, 17]. Tadqiqotlar davomida ma'lumotlarni tayyorlash, tozalash va xususiyat ajratish bosqichlarining integratsiyasi asosida, model tanlash va o'rgatish jarayonlari sinchkov tahlil qilinib, ularning har birining tizimli sxema yaratilishiga qo'shgan hissiy aniqlandi [Medhat, 2014, 35]. Bunday yondoshuvlar, qo'lda yaratilgan leksikon va avtomatlashtirilgan xususiyat ajratish metodlarini birlashtirgan holda, tasniflash sxemasining puxtaligini oshirishga xizmat qiladi.

Yuqorida olib borilgan tahlil natijalariga ko'ra sentiment tahlil qilish jarayonining umumiy tasniflash sxemasini ishlab chiqdik. Ushbu sxema sentiment tahlilini **bosqichma-bosqich** qanday amalga oshirishni ifodalaydi

1. Ma'lumotlarni yig'ish (Dataset Preparation).
2. Ma'lumotlarni tozalash va tayyorlash (Preprocessing).
3. Xususiyatlar ajratish (Feature Engineering).
4. Modelni tanlash va o'qitish (Model Training & Selection).
5. Modelni baholash va optimallashtirish (Evaluation&Optimization).
6. Modeldan foydalanish (Deployment).

1. Ma'lumotlarni yig'ish (Dataset Preparation). Ushbu bosqichda, sentiment tahlil uchun kerakli xom ma'lumotlar turli axborot manbalaridan – ijtimoiy tarmoqlar, forumlar, bloglar, onlayn sharhlar va yangilik portallaridan – yig'iladi. Ma'lumotlar yig'ilishi davomida ularning dolzarbligi, hajmi va tasniflashga mosligi aniqlanadi, shuningdek, tasniflash algoritmlari uchun zarur bo'lgan kontekstual va lingvistik xususiyatlar inobatga olinadi. Dataset namunasi quyidagi 1-jadvalda keltirilgan.

1-jadval. O'zbek tili gaplardan iborat dataset

№	Gap
1.	Bu film juda ajoyib!
2.	Xizmat sifati meni qoniqtirmadi.
3.	Bugungi ob-havo juda yaxshi edi.
4.	Bu film vaqtini behuda ketkazish edi.
5.	Ovqat juda mazali edi.

2. Ma'lumotlarni tozalash va tayyorlash (Preprocessing). Yig'ilgan xom ma'lumotlar ustida, keraksiz belgilar, HTML teglar, reklama, spam va maxsus belgilar kabi elementlarni olib tashlash operatsiyalari bajariladi. Bundan tashqari, matnlar tokenizatsiya, normalizatsiya, stop-words'larni chiqarib tashlash, stemming yoki lemmatizatsiya kabi bosqichlardan o'tkazilib, yagona va strukturallashgan formatga keltiriladi. Bunda asosan quyidagi amallar bajariladi:

- Belgilarni olib tashlash (emoji, maxsus belgilar, HTML teglar)
- Katta-kichik harflarni normallashtirish
- Nomuhim so'zlarni olib tashlash (masalan, "va", "ham", "edi")
- Tokenlash: Gaplarni alohida so'zlarga ajratish
- Lemmalash/Stemlash: So'zlarni asos shakliga keltirish

Original matn: "Bu film juda ajoyib! 🍿 #SuperFilm"

Tozalangan matn: "film juda ajoyib"

3. Xususiyatlar ajratish (Feature Engineering). Xususiyat ajratishning to'g'ri bajarilishi modelning o'rgatilishida va keyingi sinov bosqichlarida yuqori aniqlikka erishish uchun zarur

bo'lgan ma'lumotlar to'plamini shakllantiradi. Matnni **raqamli formatga o'tkazish** uchun quyidagi usullar ishlatiladi:

- **Bag of Words (BoW)** – matndagi so'zlarning chastotasi aniqlanadi.
- **TF-IDF (Term Frequency - Inverse Document Frequency)** – matndagi muhim so'zlar aniqlanadi.
- **Word2Vec / FastText / GloVe** – so'zlar vektor ko'rinishga o'tkaziladi.
- **BERT yoki transformer asosida embeddings** – Kontekst hisobga olinadi.

4. Modellarni o'qitish va tanlash (Model Training & Selection). Xususiyatlar ajratilgan dataset asosida sentiment tahlil uchun eng mos model aniqlanadi. Bu bosqichda, turli mashinali o'qitish algoritmlari (masalan, logistik regressiya, SVM, neyron tarmoqlar, va chuqur o'rganish modellaridan biri) sinovdan o'tkazilib, ularning afzalliklari va cheklovlari solishtiriladi. Tanlangan model trening to'plami yordamida o'rgatiladi, bu jarayonda gradient tushishi, kross-validatsiya va boshqa optimallashtirish strategiyalari qo'llanilib, model parametrlarining eng yaxshi kombinatsiyasi topiladi. Ma'lumotlarni o'qitish va baholash uchun **trening (o'quv), validatsiya (baholash) va test** qismlariga ajratish lozim:

- **Trening (70%)** – modelni o'qitish uchun dataset;
- **Validatsiya (15%)** – model parametrlarini moslashtirish uchun dataset;
- **Test (15%)** – modelni baholash uchun dataset.

Klassik mashinali o'rgatish (ML) modellari quyidagilar:

- **Naïve Bayes** – tez va oddiy model;
- **Logistik regressiya** – matnlarni tasniflash uchun samarali model;
- **SVM (Support Vector Machine)** – murakkab tuzilmali matnlarni yaxshi ajratuvchi model.

Chuqur o'rganish (Deep Learning) modellari quyidagilar:

- **LSTM / BiLSTM / GRU** – kontekstni tushunadigan model;
- **CNN (Convolutional Neural Network)** – matnlardagi asosiy xususiyatlarni aniqlaydigan model;
- **Transformer modellar (BERT, XLM-RoBERTa)** – Eng zamonaviy sentiment tahlil modellaridan biri.

5. Modelni baholash va optimallashtirish (Evaluation&Optimization). aholash natijalariga asoslanib, model parametrlarida qo'shimcha optimallashtirishlar, hiperparametr tuning va xatoliklarni kamaytirish bo'yicha strategiyalar ishlab chiqiladi. Ushbu bosqich, modelning umumiy lashuv qobiliyatini mustahkamlash va real sharoitlarda yuqori natijalar berishini ta'minlashda hal qiluvchi rol o'ynaydi. Modelni baholash uchun quyidagi ko'rsatkichlardan foydalaniladi:

- **Aniqlik (Accuracy)** – to'g'ri bashoratli matnlar ulushini ifodalovchi ko'rsatkich;
- **F1-score** – muvozanatlangan ko'rsatkich (Precision va Recall asosida hisoblanadi);
- **AUC-ROC** – modelning sentimentlarni ajratish qobiliyatini baholadigan ko'rsatkich.

№	Model	Accuracy	Precision	Recall	F1-score
1	Naïve Bayes	80%	78%	81%	79%
2	LSTM	85%	84%	86%	85%
3	BERT	92%	91%	93%	92%

6. Modeldan foydalanish (Deployment). Optimallashtirilgan va baholangan model real vaqtda sentiment tahlilini amalga oshirish uchun tizimga integratsiya qilinadi. Modelni joriy qilish natijasida, real axborot tizimlarida foydalanuvchilarning hissiy holatini aniqlash, sharhlar va ijtimoiy tarmoq ma'lumotlarini avtomatlashtirilgan tarzda tahlil qilish imkoniyati yaratiladi. Modelni real hayotda ishlatish uchun **API** yoki veb interfeys shakllantiriladi. Ushbu dastur interfeysi quyidagi ketma-ketlikda ishlaydi:

1. Foydalanuvchi matn kiritadi.
2. Model matnni tozalaydi va sentimentni baholaydi.

3. **Natija:** Ijobiy, salbiy yoki neytral kabi chiqaraladi.
4. **Misol:**

Kiritish: "Bu restoran xizmati a'lo darajada!"

Chiqish: "Xizmat: a'lo darajada)" → Positive

Umuman olganda sentiment tahlil qilish modelining sxemasi quyidagicha:

1-variant. Ma'lumotlarni yig'ish → Matnlarni tozalash → Xususiyatlarni ajratish → Modelni o'qitish → Modelni baholash → Modeldan foydalanish.

2 -variant. Ma'lumotlarni yig'ish → Matnlarni tozalash → Tokenlash → Model orqali sentiment aniqlash → Aspekt va opinionlarni ajratish → Natijalarni saqlash va tahlil qilish.

Xulosa. Tadqiqot natijalari, mavjud adabiyotlar sharhi va ilgari olib borilgan ilmiy ishlarga asoslanib, leksikon asosidagi va mashinali o'qitish metodlarining qo'llanilishi orqali yuqori aniqlikka erishish imkoniyatini ko'rsatdi.

Foydalanilgan adabiyotlar:

1. Cambria, E., Li, Y., Xing, F. Z., Poria, S., & Kwok, K. (2020, October). SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In *Proceedings of the 29th ACM international conference on information & knowledge management* (pp. 105-114).
2. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.
3. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1-2), 1-135.
4. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013, October). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1631-1642).
5. Tang, D., Wei, F., Qin, B., Yang, N., Liu, T., & Zhou, M. (2015). Sentiment embeddings with applications to sentiment analysis. *IEEE transactions on knowledge and data Engineering*, 28(2), 496-509.

QORAQALPOQ TILI ELEKTRON KORPUSINI YARATISH: NAZARIY, AMALIY VA STRATEGIK ASOSLAR

Abduraxmonova Nilufar Zaynobiddinovna

Filologiya fanlari doktori, professor
Mirzo Ulugʻbek nomidagi Oʻzbekiston milliy universiteti
E-mail: n.abdurakhmonova@nuu.uz

Allaberdiyeva Durdona Gurbanmurat qizi

Mirzo Ulugʻbek nomidagi Oʻzbekiston milliy universiteti magistranti
E-mail: durdonallaberdiyeva39@gmail.com

Annotatsiya: Ushbu maqolada qoraqalpoq tili uchun elektron korpus yaratishning nazariy asoslari, zamonaviy lingvistikadagi oʻrni, amaliy bosqichlari va istiqbolli yoʻnalishlari atroflicha yoritiladi. Korpus lingvistikasining muhim jihatlari, xorijiy va mahalliy tajribalar, shuningdek qoraqalpoq tili uchun bunday resursning ahamiyati va zarurligini ilmiy asosda tahlil qiladi. Shuningdek, mavjud texnologik yechimlar, resurslar va ularni bir butun korpus tizimiga birlashtirish mexanizmlari koʻrib chiqiladi.

Kalit soʻzlar: qoraqalpoq tili, elektron korpus, korpus lingvistika, til resurslari, tilshunoslik texnologiyalari, tilni raqamlashtirish, til siyosati.

Abstract: This article provides a detailed overview of the theoretical foundations of the creation of an electronic corpus for the Karakalpak language, its role in modern linguistics, practical stages and promising directions. It analyzes the important aspects of corpus linguistics, foreign and local experiences, as well as the importance and necessity of such a resource for the Karakalpak language on a scientific basis. It also considers existing technological solutions, resources and mechanisms for integrating them into a single corpus system.

Keywords: Karakalpak language, electronic corpus, corpus linguistics, language resources, linguistic technologies, language digitization, language policy.

Kirish

Korpus lingvistika bugungi kunda tilshunoslik fanining ilgʻor va zamonaviy yoʻnalishlaridan biri sifatida shakllanib ulgurgan. Uning asosiy tamoyili til hodisalarini real, empirik maʼlumotlar asosida oʻrganishdan iborat boʻlib, bu tilga oid nazariy xulosalarni ishonchli, tizimli va statistik asosda shakllantirish imkonini beradi. Bugungi globallashuv va raqamli texnologiyalar asrida har bir milliy tilning raqamli muhitga integratsiyalashuvi muhim ilmiy, amaliy va siyosiy masalaga aylangan.

Qoraqalpoq tili ham oʻzining tabiiy taraqqiyotida zamonaviy texnologiyalar bilan uygʻunlashishi, til resurslari tizimida oʻz oʻrnini egallashi va ilgʻor tilshunoslik amaliyotlarida ishtirok etishi uchun elektron korpus shaklida ifodalangan ilmiy-tadqiqot bazasiga ehtiyoj sezmoqda. Bu ehtiyoj tilni oʻrganish, saqlash, rivojlantirish va uni sunʼiy intellektga asoslangan tizimlarga kiritish nuqtai nazaridan strategik ahamiyatga ega.

Elektron korpus tilshunoslikda muhim manba sifatida eʼtirof etiladi. U tilning haqiqiy, amaliy qoʻllanish holatlaridan iborat katta hajmdagi matnlar majmuasi boʻlib, ularni maxsus formatda toʻplab, elektron vositalar orqali tahlil qilish imkonini beradi. Elektron korpuslar kompyuter lingvistikasi, sunʼiy intellekt, mashinali tarjima, avtomatik matn tahlili, leksikografiya kabi koʻplab yoʻnalishlarda keng qoʻllaniladi...

Qoraqalpoq tili turkiy tillar oilasiga mansub, mustaqil grammatik tuzilishga ega boʻlgan til boʻlib, oʻzining boy leksikasi, fonetik va morfologik tizimi bilan ajralib turadi. Biroq, bugungi kungacha qoraqalpoq tilining toʻlaqonli elektron korpusi shakllantirilmagan...

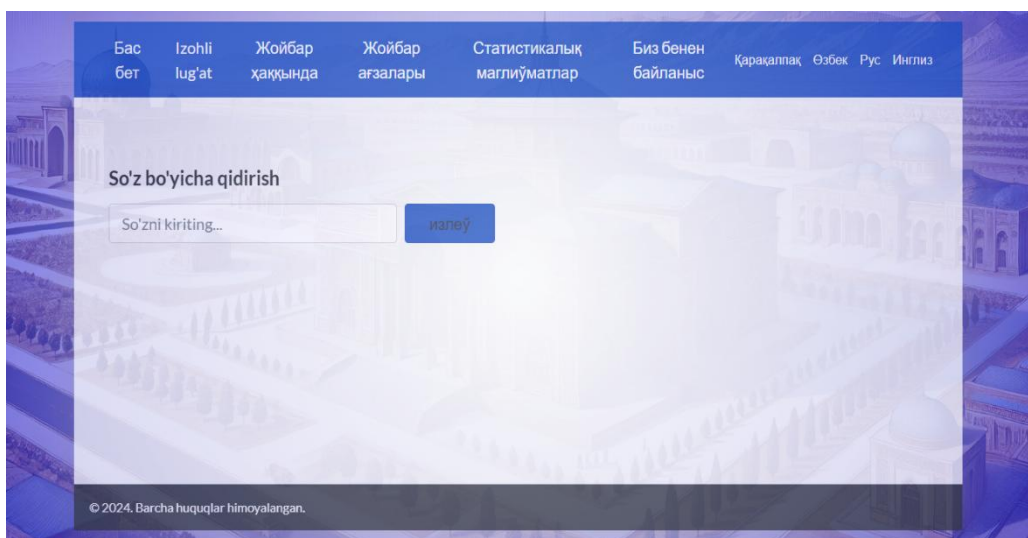
Elektron korpus yaratish ilmiy va texnologik jihatdan bir nechta o'zaro bog'liq bosqichlardan iborat. Dastlabki bosqichda turli janrdagi yozma va og'zaki matnlar yig'iladi...

Zamonaviy lingvistik korpuslar nafaqat katta hajmdagi matnlar to'plami, balki ularni semantik, sintaktik va pragmatik darajalarda avtomatik yoki yarim-avtomatik tahlil qilish imkonini beruvchi dasturiy vositalar majmuasidir. Yaratilgan Qoraqalpoq tili elektron korpusi ham ushbu tamoyillarga asoslangan holda ishlab chiqilgan bo'lib, uning interfeysi foydalanuvchilarga ilmiy va amaliy jihatdan yuqori darajadagi lingvistik tahlillarni o'tkazish imkonini beradi.

Qoraqalpoq tili elektron korpusining muhim komponentlaridan biri bu **qidiruv tizimidir**. Qidiruv interfeysi foydalanuvchiga:

- ✓ **So'z yoki ibora bo'yicha matnlarda qidirish;**
- ✓ **Leksik shakl (lemmatizatsiyalangan holda) bo'yicha izlash;**
- ✓ **Morfologik belgilarga asoslangan qidiruv** (masalan, fe'l zamoni, shaxs-son ko'rsatkichlari, ot tuslanishi);
- ✓ **Sintaktik konstruksiyalarni aniqlash;**
- ✓ **Janr yoki uslubiy ko'rsatkichlar bo'yicha filtrlash** imkonini beradi.

Qidiruv natijalari **konkordans oynasi** ko'rinishida beriladi — bu foydalanuvchiga tanlangan birlikning kontekstini yaqin atrofdagi gaplar bilan ko'rish imkonini yaratadi. Bu esa tildagi birliklarning semantik xususiyatlarini aniqlashda, kollokatsiyalarni tahlil qilishda va til birliklarining real qo'llanilishini chuqurroq anglashda yordam beradi.



Shuningdek, qidiruv tizimida **statistik tahlil** funksiyasi ham mavjud bo'lib, u so'z chastotasi, tekstdagi takrorlanish darajasi, so'z turkumlari soni kabi axborotlarni vizual shaklda (diagramma, jadval, grafik) taqdim etadi.

Interfeysga integratsiyalashgan izohli lug'at — korpusdagi asosiy leksik resurslardan biri hisoblanadi. Ushbu lug'atda qoraqalpoq tilidagi leksik birliklar morfologik, semantik va stilistik jihatdan batafsil tasniflangan. Har bir so'z shakli uchun quyidagi ma'lumotlar taqdim etiladi:

- ✓ **Asosiy ma'no (denotativ ma'no);**
- ✓ **Qo'shimcha ma'no va konnotatsiyalar;**
- ✓ **Uslubiy belgi (rasmiy, og'zaki, badiiy, ilmiy);**
- ✓ **Sinonimik va antonimik munosabatlar;**
- ✓ **Korpusdan olingan real kontekstlardagi qo'llanmalari.**

Bu lug'at korpus matnlari asosida avtomatik tarzda boyitilib boradi, ya'ni foydalanuvchilar kontekstli ma'nolarni real foydalanish doirasida kuzatish imkoniyatiga ega. Bu, o'z navbatida, lingvodidaktik tadqiqotlar, tarjima amaliyoti va mashina o'rgatish modellarini yaratishda muhim ahamiyat kasb etadi.

Бас бет	Izohli lug'at	Жойбар ҳаққида	Жойбар ағзалары	Статистикалық мағлұматлар	Биз бенен байланыс	Қарақалпақ	Өзбек	Рус	Инглиз
	so'z	Turkum	Izoh						
	Abadan	Kelbetlik	Jetkilikli, kóp, qurğın, barshlıq, Zamanimiz endi abadan. Kámbağalğa náwbet jetip basladi (A.Muwsaev). Biz paxtatesh xalıqlar azat shıgısta, Jasaymız abadan shadlı turmısta (I.Yusupov). ádıra qalğır tuwğın jer, Bolmağın soñ abadan (Jiyen Jiraw).						
	Abadanlasıw	Feyil	Qurğınlasıw molshılıqqa erisiw, bayıw, güllenıw, ósip jetisiw. Bizın maqsetimiz kúnnen kúnge abadanlasıw.						
	Abadanlastırıw	Feyil	Abadanlasıw feyilinin' o'zgelik da'rejesi						
	Abadanlıq	Atlıq	Molshılıq, toqshılıq, barlıq, qurğınlıq, barlıq, Quwanish qoyınğa tolğın dáwirde, Baxit abadanlıq bolğın dáwirde (I. Yusupov). Abadanlıq hám güllengen dáwirde, Xalqımızdın qálep alğın erki bul (I.Yusupov).						
	Abadanshılıq	Atlıq	Molshılıq, toqshılıq, barlıq, qurğın dáwir, abadanlıq, Arıq mergen jigiteri menen añ awlap, eli, xalqı abadanshılıqta jasap, jaqsı turmıs keshiripti (Q. X.E).						
	Abajur	Atlıq	Janlıqtı bir orınğa jıynaqlap toplaw ushin lampanıń ústine ornattılğın (gezleme, qağaz tağı basqa nárselerden jasalğın) xojalıq úskinesı, shıra qalpaq.						

Korpus matnlari **metama'lumotlar bilan belgilangan** bo'lib, bu korpusdagi har bir matn haqidagi texnik, lingvistik va sotsiolingvistik axborotlarni o'z ichiga oladi. Har bir matnga quyidagi metama'lumotlar biriktirilgan:

- ✓ **Matnning manbaasi (gazeta, sayt, kitob, nutq va h.k.);**
- ✓ **Yozilgan yoki e'lon qilingan sanasi;**
- ✓ **Muallif ismi yoki anonimlik darajasi;**
- ✓ **Janri (badiiy, ilmiy, publitsistik, rasmiy va h.k.);**
- ✓ **Topik (mavzu): siyosat, madaniyat, ta'lim, sport, va boshqalar;**
- ✓ **Yozuv turi (lotin yoki kirill);**
- ✓ **Dialekt yoki mintaqaviy variantlar (mavjud bo'lsa).**

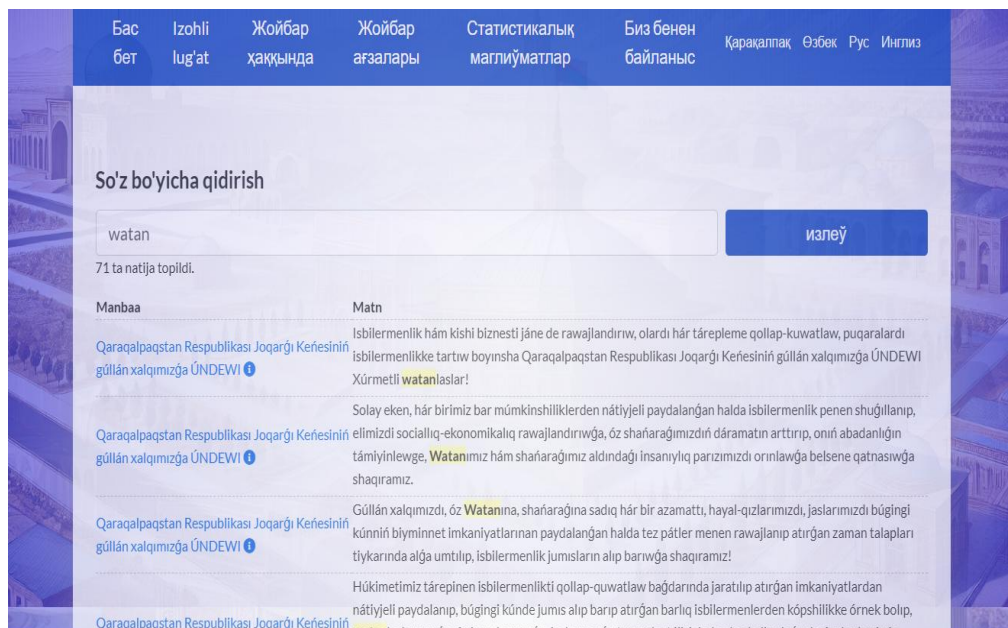
Mazkur metama'lumotlar foydalanuvchiga matnlar ustida chuqur qiyosiy tahlil o'tkazish imkonini beradi. Misol uchun, biror leksik birlikning qanday janrlarda yoki davrda ko'proq ishlatilishini aniqlash, til evolyutsiyasini o'rganish, matnlarning sotsiolingvistik xarakteristikasini ochib berish mumkin bo'ladi.

Qoraqalpoq tili elektron korpusi interfeysi o'zida nafaqat korpus matnlarini jamlagan, balki ularni tahlil qilish, izlash, tasniflash va semantik tafsilotlarga ega bo'lgan mukammal lingvistik platformani mujassamlashtirgan. Izohli lug'at, qidiruv mexanizmi va metama'lumotlar tizimi orqali foydalanuvchilar tilshunoslikning barcha darajalarida — leksikologiya, sintaksis, stilistika, pragmatika — bo'yicha chuqur tahlil olib borishlari mumkin. Bu esa qoraqalpoq tili uchun birinchi yirik raqamli til resursining shakllanishini anglatadi. Ushbu dissertatsiyada ilgari surilgan nazariy va amaliy g'oyalar qoraqalpoq tilining zamonaviy korpusini yaratishga xizmat qiladigan tayanch bilimlar majmuasidir. Amalga oshirilgan ishlar tilshunoslik fanining rivojlanishiga, qoraqalpoq tilining raqamli taraqqiyotiga va o'quv-metodik, tarjima, sun'iy intellekt sohalaridagi qo'llanilish imkoniyatlarini kengaytirishga asos bo'ladi.

Shu bilan birga, mazkur izlanishlar natijasida quyidagi ilmiy-amaliy xulosalarga kelindi:

- ✓ Qoraqalpoq tilining zamonaviy elektron korpusini yaratish uchun nazariy, texnologik va amaliy asoslar mavjud bo'lib, ularni integratsiyalash orqali raqamli til resursi shakllanishi mumkin.
- ✓ Turkiy tillarning korpus lingvistikasi tajribasi qoraqalpoq til korpusi modelini ishlab chiqishda metodologik asos bo'lib xizmat qiladi.
- ✓ Matnlar klassifikatsiyasi, lug'at tuzish, metama'lumot yaratish, tokenizatsiya va foydalanuvchi interfeysi – bu korpus yaratishning ajralmas bosqichlaridir.
- ✓ Lingvistik korpuslar tilshunoslikdan tashqari boshqa sohalar – tarjima, ta'lim, sun'iy intellekt, tilni o'rgatish, lingvodidaktika, ijtimoiy tarmoqlarni tahlil qilish kabi yo'nalishlar uchun ham beqiyos manba hisoblanadi.

- ✓ Dissertatsiyada ishlab chiqilgan tavsiyalar asosida qoraqalpoq tili korpusining dastlabki elektron platformasi yaratildi.



Xulosa qilib aytganda, qoraqalpoq tili elektron korpusining yaratilishi – bu korpus tilshunoslik fanida yangi bosqich bo‘lib, nafaqat nazariy, balki amaliy ahamiyatga ham egadir. U til taraqqiyoti, xalqaro integratsiya, zamonaviy ta‘lim vositalari, axborot texnologiyalari bilan uyg‘unlashgan holatda tilning ijtimoiy hayotdagi mavqeini mustahkamlashga xizmat qiladi.

Foydalanilgan adabiyotlar:

1. McCarthy M. *Discourse Analysis for Language Teachers*. – CUP, 1991.
2. McEnery T., Wilson A. *Corpus Linguistics*. – Edinburgh Univ. Press, 1996.
3. McEnery, Tony, and Andrew Wilson. *Corpus Linguistics: An Introduction*. Edinburgh University Press, 2001
4. Meyer C. *English Corpus Linguistics: An Introduction*. – CUP, 2002.
5. Partington A. *Patterns and Meanings*. – Amsterdam: Benjamins, 1998.
6. Reppen R. *Using Corpora in the Language Classroom*. – CUP, 2010.
7. Sinclair J. *Corpus, Concordance, Collocation*. – Oxford Univ. Press, 1991.
8. Stefanowitsch A. *Corpus Linguistics: A Guide to the Methodology*. – de Gruyter, 2020.
9. Abdurakhmonova, N., Shakirovich, I. A., Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, 2022. - 146-156b.
10. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., Abduvahobo, G. UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. *TURKLANG 2022*, -231-240b.
11. Uzbek Electron Corpus: uzbekcorpus.uz. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.
12. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE.

O'ZBEK VA TURK TILLARIDA JOY NOMLARINING IFODALANISHI

(Representation of place names in Uzbek and Turkish languages)

Turayeva Umida Toxirjon qizi

Mirzo Ulug‘bek nomidagi

O'zbekiston Milliy universiteti magistranti

E-mail: umidasng@gmail.com

Annotatsiya: Mazkur maqolada o‘zbek va turk tillarida mavjud joy nomlarini lingvistik-qiyosiy tahlili chog‘ida ularning umumiy va farqli jihatlari aniqlanib, semantik-struktural va etimologik tahlil qilindi. Shuningdek, ikki tildagi joy nomlarini qiyoslab, ularning etnik kelib chiqishi va ifodalanishiga oid misollar keltirib o‘tildi. Toponimika, jumladan, o‘zbek toponimikasi tarix, etnografiya, genealogiya, geraldika, matnshunoslik, adabiyotshunoslik, geografiya, astronomiya, geologiya, demografiya kabi bir qator fanlar bilan uzviy aloqador fan sanaladi. Ayniqsa, bu sohaning antroponimika, toponimika, etnonimika kabi bir qator sohalarida juda ko‘plab ishlar amalga oshirilgan. Maqolada toponimika borasida ilmiy tadqiqot olib borgan tadqiqotchilar haqida ham fikr yuritilgan.

Kalit soʻzlar: toponimika, etnografiya, qiyosiy tahlil, etimologiya

Abstract: This article presents a linguistic-comparative analysis of toponyms in the Uzbek and Turkish languages, identifying their common and distinct features, and providing semantic-structural and etymological analyses. Additionally, place names in both languages were compared, with examples given of their ethnic origins and expressions. Toponymy, including Uzbek toponymy, is considered a science closely interconnected with various disciplines such as history, ethnography, genealogy, heraldry, textual studies, literary studies, geography, astronomy, geology, and demography. Notably, extensive work has been carried out in several subfields of this area, including anthroponymy, toponymy, and ethnonymy. The article also discusses researchers who have conducted scientific studies in the field of toponymy.

Keywords: toponymy, ethnography, comparative analysis, etymology

Аннотация: В данной статье при лингвистическом сравнительном анализе топонимов, существующих в узбекском и турецком языках, выявлены их общие и отличительные черты, проведен семантико-структурный и этимологический анализ. Также были сопоставлены топонимы на двух языках и приведены примеры их этнического происхождения и способов выражения. Топонимика, в том числе узбекская топонимика, считается наукой, неразрывно связанной с рядом дисциплин, таких как история, этнография, генеалогия, геральдика, текстология, литературоведение, география, астрономия, геология, демография. В частности, в этой области проведено множество исследований в таких направлениях, как антропонимика, топонимика, этнонимика. В статье также рассматриваются работы исследователей, проводивших научные изыскания в области топонимики.

Kirish

Hozirda o'zbek va turk tillaridagi so'zlarni, ayniqsa, joy nomlarini har turli jihatdan o'rganishga e'tibor kuchayib bormoqda. Ko'pincha toponimlar grammatik va semantik jihatdan hozirgi shaklga kelgunga qadar talay o'zgarishlarga uchraydi. Toponimlarni paydo bo'lishiga qarab ikki turga bo'lish mumkin: tabiiy yo'l bilan hosil bo'lgan, xalq tomonidan qo'yilgan toponimlar va sun'iy yo'l bilan hosil qilingan toponimlar. Bundan kelib chiqqan holda, ikki tilda ham o'ziga xos ajoyib toponimlar mavjud, ikki tilning ham ildizi uzviy bog'liq bo'lib, ba'zi so'zlar va joy nomlari ham shakliy, ham ma'no jihatidan juda o'xshashdir. Ushbu maqolamizda

toponimlarning ayni shu jihatiga, ya'ni ikki tildagi bir-biriga o'xshash bo'lgan joy nomlariga ba'zi misollarni keltirib, ularning semantik tahliliga e'tibor qaratmoqchimiz.

Adabiyotlar tahlili va metodologiya

Insoniyat yaralibdiki, atrofini o'rab turgan har qanday belgi-predmetni nomlashga shu orqali muloqotning oddiy va samarali yo'llarini yaratishni maqsad qilgan. Keyinchalik bu an'ana davom etib, fanda o'z o'rganish obyekti va predmetiga ega nomshunoslik-onomasiologiyaning vujudga kelishiga sabab bo'ldi. Umuman, insonlar narsalarga nom qo'ymasdan yashay olmaydilar, dunyodagi hamma narsaning nomi bor. O'simlik va hayvonlarning mahalliy nomlaridan tashqari ilmiy, lotincha nomlari ham mavjud. Har qanday o'simlik va hayvonni lotincha nomiga qarab turli millat mutaxassisilari ilmiy faktik nazariy manba va materiallarni almashganlar. Demak, nomlar bir narsani ikkinchi narsadan farq qilishga yordam beradi. Bir joyni ikkinchisidan, bir jilg'ani boshqa bir jilg'adan, bir ko'chani yonidagi ko'chadan, tog'-adirlarni, shahar va qishloqlarni, bir-birlaridan farq qilish uchun ham kishilar nomlar o'ylab chiqargan. Joy nomlari, geografik nomlarni tilshunoslikda onomastikaning bir bo'limi **toponimika** fani o'rganadi. Joy nomlari, ya'ni toponimiya bir necha turga bo'linadi. Bular gidronomiya (yunoncha gidro - suv), ya'ni daryolar, ko'llar, dengizlar, soylar, kanallar, qo'ltiqlar, bo'g'ozlar, sharsharalar nomlari. Oronimiya (yunoncha oros — tog'), ya'ni yer yuzasining relyef shakllari - tog'lar, cho'qqilar, qirlar, vodiylar, tekisliklar nomlari; oykonimiya (yunoncha oykos - uy), polinimiya (yunoncha polis — shahar) yoki urbanomiya (lotincha urbos — shahar), ya'ni qishloq shaharlarning nomlari, mikrotoponimiya (yunoncha mikros -kichik), ya'ni kichik obyektlar: buloqlar, quduqlar, dalalar, o'tloqlar daraxtzorlar, jarlar, yo'llar, ko'priklar va hatto atoqli otga ega bo'lgan ayrim daraxtlar nomlari. Bundan tashqari, turli xalq, urug'-aymoq nomlari bilan atalgan toponimlar etnotoponimlar (yunoncha "etnos" - xalq) deyiladi. Kishi ismlari bilan yuritiladigan toponimlarni esa antropotoponimlar (yunoncha "antropos" - odam) deb atash mumkin. Shu kunga qadar o'zbek toponimikasi bo'yicha 20 dan ortiq nomzodlik, uchta doktorlik (Z.Do'simov, S.Qorayev, T.Enazarov) ishi himoya qilindi. O'zbek toponimikasi bo'yicha himoya qilingan nomzodlik ishlari Qashqadaryo viloyati (T.Nafasov, T.Enazarov, O.Begimov), Surxondaryo viloyati (Y.Xo'jamberdiev, X.Xolmo'minov), Samarqand viloyati (T.Rahmatov, N.Begaliev, A.Turobov, S.Bo'riev), Sirdaryo viloyati (N.Mingboev), Toshkent viloyati (M.N.Ramazonova), Buxoro viloyati (S.Nayimov, A.Aslonov), Navoiy viloyati (O'.Oripov), Jizzax viloyati (A.Zokirov), Farg'ona viloyati (N.Oxunov, S.Gubaeva, J.Latipov), Xorazm viloyati (Z.Do'simov, A.Otajonova, M.Tillaeva), Shimoliy o'zbek shevalari toponimiyasi (L.Karimova), Toshkent shahri (Sh.Qodirova) toponimiyasi bo'yicha yozilgan

Toponimlar asosan apellyativlar asosida hosil bo'ladi. *Apellyativ* lotincha *appellativum* – *turdosh ot* so'zidan olingan bo'lib, atoqli otga qarama-qarshi qo'yiluvchi, atoqli ot yasash uchun lisoniy baza bo'luvchi turdosh ot va boshqa til leksemalaridir.

Toponimlarda xalqning o'tmishi, urf-odatlar, boshidan kechirgan turli xil siyosiy-ijtimoiy, etnomadaniy jarayonlar o'z aksini topgan bo'ladi. Shunga binoan toponimiya muayyan til lug'at boyligining muhim tarkibiy qismi, o'tmish voqealari darakchisi sanaladi.

Toponim – geografik nom ham so'z. Lekin toponimlar atoqli otlardir. Atoqli otlar til taraqqiyotining qiyosan keyingi bosqichlarida turdosh otlar (apellyativlar)dan kelib chiqqan. Qadimiy tillarda atoqli otlar bo'lmagan deyishadi. Avstraliya, Afrika va Amerikaning tub joy xalqlari — aborigenlar tillarida atoqli otlar juda kam ekan. Bunday tillarda turdosh otlar, biron so'z birikmalari atoqli otlar vazifasini bajarar ekan. Xuddi shunday so'z birikmalari tobora turg'un shakl oladi. Odamlar tevarak-atrofidagi o'zlariga tanish kichik hududdagi joylarni «Baliq tutiladigan ko'l», «Sel keladigan soylik», «Qo'y qirilgan yaylov», «Cho'chqali soy» kabi butun bir gaplardan iborat so'z birikmalari bilan ataydilar. Konkret geografik obyektlarni ifodalaydigan bunday so'z birikmalari va turdosh otlar ko'p bo'lmaganidan bora-bora atoqli otlarga aylanadi. Turdosh otlar ana shu yo'l bilan toponimlashadi, ya'ni toponimlarga aylanadi. Masalan, olma so'zi umuman olma bog'larini ifodalaydigan umumiy tushuncha ekan, turdosh otligicha qolaveradi. Bu tushuncha individuallashtirish, konkret bir obyektning ifodalay boshlasa va shunday qilib ayrim bir

tushunchani bildirsa, atoqli otga aylanadi qoladi. Katta jar soʻzlari individuallashmas, konkretlashmas, yaʼni biron obyektini ifodalamas ekan, salbiy relyef shakllarini bildiradigan turdosh ot, apellyativ boʻlib qolaveradi. Shunday qilib, toponimlar geografik tushunchalarning konkretlashishi hamda individuallashishi jarayonida turdosh otlardan paydo boʻlgan va paydo boʻlaveradi.

Toponimikaga oid koʻplab dasturlar, metodik vositalar va dars vositalari bor. Ammo Y. B. Yusufov va S. K. Kerimovning "Toponimika asoslari" kitobi Ozarbayjon tilida darslik tayyorlash sohasidagi birinchi namunadir. Ushbu darslik Ozarbayjon toponimiyasiga sharh boʻlib, ularning yutuqlarining asosidir. Kitoblarda mualliflar oʻz toponimik gʻoyalarini ilgari surdilar va toponimikaning barcha dolzarb masalalarini yangicha nuqtayi nazardan izohlashga harakat qildilar. Turkiyada toponimlarning oʻrganilish tarixi bilan bir qatorda boshqa mamlakatlarda oʻtkazilgan toponimik tekshirishlar ham yuqori baholandi. "SSSR toponimiyasi: Toponimikaning umumiy xususiyatlari" deb nomlangan, beshinchi bobda keltirilgan maʼlumotlarning bir boʻlimi V.A. Junchkevichning "Obshaya toponimi" kitobidan olingan boʻlsa-da, boʻlim tanqidiy shaklda qayta koʻrib chiqildi. Ilovalar rahbarligi ostida - maqola boʻyicha taqdim etilgan materiallar amaliy ahamiyatga egadir. Toponimlarning umumiy dolzarb muammolari izohlansa-da, misollar asosan Ozarbayjon toponimiyasidan berilgani uchun talabalarga tushunarli holatdadir. Professor M. A. Gurbanov shahar nomlari, shuningdek, ular kelib chiqqan madaniyat haqida koʻplab sirlarni oʻz ichiga oladi. Asrlar davomida turli sivilizatsiyalar uyi boʻlgan Turkiya hozirgi kungacha gullab-yashnagan sivilizatsiyalarning izlarini tarixiy qoldiqlar bilan ham, shahar nomlari bilan ham olib yurishgan. Turkiyadagi shahar nomlari Istanbuldan Trabzonga, Izmirdan Amasiyaga qayerdan kelganini tahlil qilsak.

Istanbul hozirgi nomini olishdan oldin, u turli nomlar bilan atalgan. Miloddan avvalgi 658 yilda Megara qiroli Byzas tomonidan tashkil etilganida, u Vizantiya nomi bilan atalgan. Rim imperiyasi hukmronligi ostida boʻlganida, Rim imperatori Markus Avreliy davrida u imperatorning majoziy otasi nomi bilan "Antion" deb nomlangan. Keyinchalik, yaʼni Vizantiya davrida "Konstantin yoki Konstantinopol" deb atalgan. Sharq madaniyatida uni "Kostantiniye", rimliklar tomonidan esa "Konstantinopol" deb atashgan. Mahalliy aholi shahar markazi degan maʼnoni anglatuvchi "Stin-Polis" qisqartmasidan foydalangan. Biz hozirda foydalanayotgan Istanbul nomi, mahalliy aholi tomonidan ishlatiladigan ushbu qisqartma vaqt oʻtishi bilan oʻzgarganligidan kelib chiqadi.

Qadimgi manbalarga koʻra, Vizantiya tarixchisi Istefan **Adanaga** Tarsus bilan jang qilgan Adanos va Saros ismli ikki aka-uka asos solgan deb yozgan va Adanos oʻz nomini shaharga, ukasi Saros esa daryoga bergan deb keltirgan. Garchi bu rivoyat mifologik voqeaga oʻxshasa-da, Adana tarixi juda uzoq vaqtdan boshlangan turar-joy maydoni ekanligi bilan ajralib turadi. Xetlar butun Chukurova Uru Adaniya (Adana viloyati) deb ataganlar. M. U. 1650 yilda xet va Kizzuvatna qirollari oʻrtasida tuzilgan shartnomada Dekaniya nomi ishlatilgan. Qoratepa bitiklarida Danuna iborasi tilga olinadi.

Gomer dostonlariga koʻra, shahar oʻz nomini Kipr qiroli Kinyrasning qizi Smirnodan olgan. Shunga qaramay, mifologiyaga koʻra, maʼbuda Artemis Izmirdan. Qadimgi shahar bugungi kunda Konak tumani chegaralarida joylashgan arxeologik joy. **Trabzon** oʻziga xos tembrini eng koʻp saqlaydigan shahar nomlaridan biridir. Trabzon nomi Yunonistonning Pontus shtatida ishlatiladigan shaharning asl nomi boʻlgan va toʻrt burchakni anglatuvchi "Trapezus" soʻzidan kelib chiqqan. Lazlar uni Trapuzani, gruzinlar uni Trapizoni, Usmonlilar va forslar esa shaharni Tara Boʻzan deb atashgan. Ism talaffuzda biroz farq qilsa-da, hozirgi kunga kelib Trabzon Pontik davridagi nomi bilan ataladi.

Insoniyat qadimgi madaniyatining jahonga mashhur oʻchoqlari – Bobil, Afina va Rim kabi Samarqand ham koʻp asrlik boy tarixga ega. Samarqandning tarixi jahon tarixining tarkibiy qismidir. Oʻtgan 25 asr davomida shahar aholisi koʻp voqea-hodisalarni boshidan kechirdi. Samarqand Iskandar Makedonskiyning sonsiz suvoriylarini koʻrdi, arablar bosqinini boshidan kechirdi, shaharni Chingizxonning vahshiy qoʻshinlari vayron qildi. Samarqandni davlatining

poytaxti qilmoqchi bo'lgan Amir Temur o'zining ko'p sonli lashkarini shu shaharga to'plagan. Biz Samarqand tarixi haqida emas, balki uning mahalla ko'ylari to'g'risida, shahar tarixi va iqtisodiyotidan hikoya qiluvchi toponimlari haqida so'z yuritamiz. Samarqand viloyatida tarixiy yodgorliklar nomlari, Amir Temur va Temuriylar davri bunyodkorlik faoliyati bilan bog'liq joy nomlari juda ko'p. Shulardan ayrimlariga to'xtalib o'tmoqchimiz.

To'qimdo'zi – Samarqand shahridagi mahalla nomi. Bu mahalla aholisi o'tmishda to'qimdo'zlik kasbi bilan shug'ullangan.

Suzangaron – Samarqand shahridagi mahalla nomi. Mahalliy aholining o'tmishdagi asosiy kasbi yoki mashg'uloti igna tayyorlash bo'lgan. Tojik tilida Suzangaron so'zi igna tayyorlovchilar, igna ustalari degan ma'noni bildiradi.

Omonqo'ton – Samarqand viloyati Urgut tumanidagi Chaqilkalon tog'larining shimoliy yonbag'ridagi Bulbulzor soyida joylashgan g'or nomi. Bu g'or O'rta Osiyoning paleolit davridagi ibtidoiy odamlarning manzilgohi bo'lgani bilan mashhur. Bu toponimning kelib chiqishi haqida mahalliy xalq guvohlik berishicha, quyidagicha bo'lgan. O'tmishda pastki qishloqlardagi boylarning qo'ylari tog'da boqilgan. Bir yil qish juda qattiq kelib, tog'ni qalin qor qoplaydi. Bir necha kun bo'ron turadi. Toqqa borlmagan boylar qo'ylar nobud bo'ldi, deb o'ylashadi. Sovuqni zahri qaytganda toqqa kelishsa, qo'tondagi qo'ylar ham, cho'ponlar ham omon qolgan ekan. Shundan keyin bu joy nomi Omonqo'ton qishlog'i deb atalgan. G'or esa qishloq nomi bilan atalib, Omonqo'ton g'ori deb ataladi.

Ko'k-saroy – Amir Temur qarorgohi. Bu yerda asosiy davlat idoralari joylashgan edi. Saroy to'rt tosh oynali bo'lib, ko'k koshinlar bilan qoplangan. U yerda Ko'k tosh bo'lib, yirik parallelepiped shaklida ishlangan, o'yma naqshlar va guldor chiziqlar tushirilgan kulrang marmardan iboratdir.

Registon – «Qumloq joy» degan ma'noni bildiradi. Samarqanddagi mashhur maydon. Savdo va hunarmandchilik markazi. Maydonda Ulug'bek, Sherdor va Tillakori madrasalari joylashgan.

Qadimdan qishloq, mahalla va shaharlar geografik o'rni, hududning tabiiy xususiyatlari, shu yerda yashovchi aholining etnik kelib chiqishi yoki asosiy kasb-koriga ko'ra nomlab kelingan. Shuningdek, muayyan bir hududga nom berishda tarixiy, ijtimoiy-iqtisodiy va siyosiy o'zgarishlar ham hisobga olingan. Shu bois, Borsakelmas, Charchamang, Kuyganyor, Qolgan sir, Moxovtovuq, Og'zikeng, Chopon tiqqan, Gadoytopmas, Ko'kko'z, Chivintepa, Tanga topdi, Serka qirildi, Qumboyg'undi, Qulburun, Jumaloq kabi toponimlar ham o'sha yerda yashab o'tgan ajdodlardan qolgan tarixiy yodgorliklar hisoblanadi. Bu nomlar o'zida ko'plab xalq rivoyat va afsonalarining paydo bo'lishiga sabab bo'lgan lisoniy obidalaridir. Toshkent viloyatimizdagi Parkent, Piskent va Chinoz shaharlari ham xuddi shunday uzoq tarixga ega. «Sobliq», «Sofliq», «Soyliq» shaharlardan tashqari ham Toshkent vohasida yana bir necha mashhur turk toponimlari bor. Shunday toponimlardan biri hozirgi kunda Soyliq qishlog'i nomida saqlanib qolgan. Bu toponim «Sobliq» shaklida arab manbalarida zikr etiladi. Ibn Havqal asarida Shosh vohasining shimoli-sharqida joylashgan Sobliq tog'lari tilga olingan. Ushbu nom Ishtaxriyning «Masolik al-mamalik» asarida «Soflig'» tarzida qayd etiladi. Turkshunos V.V. Bartold esa Sobliqni «Soyliq» shaklida o'qish lozim, deb hisoblaydi. O'zbekistonlik olim G'aybullab Boboyorov fikriga ko'ra, bu nom dastlabki manbalarida uchraganidek, Sobliq shaklida bo'lib, so'ngra Sofliq, keyinchalik esa Soyliq shaklini olgan: Sobliq-Sofliq-Soyliq. Qadimgi turk tilida b-v-y almashuvi hodisasi bo'lgan. Masalan, qadimgi turkcha eb–ev–uy — «uy» ma'nosida, sub–suv–suy — «suv», teba–teva–tuya — «tuya» kabi. Bundan tashqari, Mahmud Qoshg'ariyning «Devonu lug'otit turk» asarida Yettisuvdagi tuxsi va chigil turkiy qabilalari yashagan shaharlardan biri «Soblyg'» shaklida uchraydi. Bu esa Shosh vohasidagi «Sobliq» va Yettisuvdagi «Soblig'» toponimlarining adash nom bo'lganini ko'rsatadi.

Aniq otlardan bir guruhi faqatgina bir borliq yoki muayyan bir guruhni ko'rsatishga xizmat qiladi: Turgut, Sarman (mushukning oti), Adana, Istanbul, Rize va h. Bular atoqli otlar (ozel adlar) dir. Atoqli otlar jonli mavjudotlar, predmet va narsalarning dunyoga kelgandan keyin olgan

xususiy nomlaridir. Atoqli otlar dunyoda yagona, o'xshashi bo'lmagan mavjudot va predmetlarni bildiruvchi so'zlar, inson, hayvon va joy nomlaridir. Boshqa bir shaklda turk tilshunoslari ularni "taqib qo'yiladigan" ma'nosida "yorliq otlari" ham deyishadi: Ahmet, Ibrahim, Erol, Turkan, Yilmaz, Tolga, Ankara, Almaniya kabi. Atoqli otlar, aslida, ma'nosiz so'zlardir. Chunki bir predmetga qarab turib, uning nomini aytish mumkin (masalan: kitap (kitob), canta (sumka), televizyon (televizor), lekin bir insonga qarab, uning ismini topib aytishning iloji yo'q. Masalan, turdosh ot bo'lgan bir meva (olma) ga qarab turib, uning qanday meva (ya'ni, olma) ekanligini oson aytish mumkin, lekin ismi Ahmet bo'lgan bir kishiga qarab turib, uning ismini aytib berishning iloji yo'q. Shu bois atoqli otlar ko'p ma'noli ham bo'la olmaydi. Agar atoqli ot turdosh otdan o'zlashgan bo'lsa, asliga ko'ra ma'noga ega bo'ladi, lekin bu o'sha atoqli otning ma'nosi emas, u o'zlashgan turdosh otning ma'nosi hisoblanadi. Masalan, Yildirm (erkak ismi) - «chaqmoq», Deniz (qiz bolaning ismi) - «dengiz», Mert (erkak ismi) - «mard», Ozgur (erkak ismi) - «ozod» va h. Turdosh otligida ma'noga ega bo'lgan bu so'zlar keyinchalik atoqli otga aylangan. Atoqli ot o'zlashtirilayotganda, uning turdosh otligidagi ma'nosi ko'zda tutilgan. Ya'ni inson o'z farzandiga Mert (mard) yoki Ozgur (ozod) ismini qo'yayotganda bu otlarning asl ma'nosidan kelib chiqqan. Atoqli otlar turdosh otlar bilan aralashib ketmasligi uchun katta harf bilan yoziladi: Oguz, Ugur, Istanbul, Turkiye, Pamuk (hayvonga qo'yilgan nom), Qahkucu (roman) va h. So'z birikmasi shaklidagi tashkilot va muassasa nomlari, turli nashr va asar nomlari tarkibidagi so'zlar har biri bosh harf bilan yoziladi: Istanbul Universitesi Edebiyat Fakultesi, Tashkent Devlet Sarksinasik Enstitusu, Hurriyet Gazetesi kabi. Shuningdek, millat, til nomlari ham atoqli otlar sifatida gapdagi o'rnidan qat'iy nazar bosh harf bilan yoziladi: Ozbek, Ozbekce kabi.

-ca (-ce, -ga, -ge). Bu qo'shimcha, aslida, shakl yasovchidir (insan-ca, yavaj-fa), ammo keyinchalik u o'z vazifasini o'zgartirib, so'z yasovchi qo'shimchaga aylangan (qarang: Tenglik shakli - Esitlik sekli). Ba'zan bu qo'shimcha qoliplashgan holda sifatlardan otlar yasashda qo'llanilgan: ala-ca, kara-ca, ak-ga, ak-ge kabi. Bunday qoliplashgan shakl ko'pincha Turkiyadagi joy nomlarida kuzatiladi: Sutlu-ce, Kanli-ca, cekme-ce, Tuzlu-ca, Derin-ce kabi.

Turk tilida joy nomlarining ba'zilarida ham kichraytirishni ifodalovchi -cik qo'shimchasi uchraydi: Ovacik, Qarcik, Galcik kabi. -cik qo'shimchasi fransuzchadan arabchaga tarjima qilingan kichraytirish otlarining o'rnini bosadi: gramue = hubeybe = tanecik, cerveret = muheyh = beyincik kabi. -q qo'shimchasi kam ishlatiladi va kuchaytiruv ma'nosini ifodalaydi. Uni faqat ana-g, ata-g so'zlarida uchratish mumkin. Quyida turk tilidagi joy nomlari va ularning etimologiyasi bilan tanishishingiz mumkin:

Qahramonmaras

Uning haqiqiy ismi Markasi. So'zlashuvda u Marashga o'zgaradi. U 11-fevral 1922-yil Mustaqillik urushida fransuzlarga qarshi o'z shahrini himoya qilish uchun qahramonlik qilib, parlament tomonidan "qahramon" unvoni bilan taqdirlandi.

Kari

Miloddan avvalgi 130-127 yillarda u yerda o'rnatilgan Karsak o'ymakorligi tufayli shahar kars deb atalgan. Kars so'zining ma'nosi tuya yoki qo'y junidan qilingan kiyim kamarni anglatadi.

Kastamonu

Shaharning eski nomi "Tumana". Keyinchalik bu yerda gaz-gaz degan qabila joylashadi. Bu yerda gaz va tumanning birlashishidan Kastamonu hosil bo'lgan.

Kayseri

Rimliklar Mazaka degan shaharni olishgach, uni Kaysarea deb atashgan. Demak, bu imperator shahri demakdir. Keyin Kayseri deb xalqda tarqalgan.

Kirsehir

Dasht va shahar so'zlarining birikuvidan hosil bo'lgan.

Kocaeli

O'rxon G'oziy davrida bu viloyatni bosib olgan Oqtuyoqo'ja ismli sarkarda tufayli uni Kocaeli deb atashgan.

Xulosa

Xulosa qilib aytganda, har bir til o'z doirasidagi so'zlarni o'rnida qo'llash, ma'no kengayish usullarini o'rganish, o'ziga xos bo'lgan xususiyatlarini dunyoning tildagi ko'rinishi orqali aks ettirishda muhim rol o'ynaydi. Shuningdek, bugungi kunda o'zbek tilida qo'llanilayotgan joy nomlari bilan turk tilida qo'llanilayotgan joy nomlari ham umumiylikni namoyon qilmoqda. Joy nomlarining nomlanish maqsadlari, ko'rib chiqish sohalari va tamoyillari baholanganda, joy nomlari alohida ahamiyatga ega hisoblanadi va uni fan sohasi sifatida o'rganish zarurligi tushuniladi. Shuning uchun keng qamrovli joy nomlari atamaları lug'atini tayyorlash lozim. Mamlakatimizda joy nomlari bo'yicha tadqiqotlar olib boradigan olimlarning ehtiyojlarini qondiradigan lug'at hali yaratilmagan. Ilgari joy nomlari bilan bog'liq bo'lgan kitoblar va maqolalarni skanerlash natijasida olingan. O'zbek va turk tilidagi joy nomlariga nazar soladigan bo'lsak, ikkala millatda ham nomlar joyning kelib-chiqish tarixi, madaniyati, urf-odatları yoki ajdodlariga asoslanib qo'yilgan. Aslida ikki xalq ham tarixdan bir urug'ga, ajdodlarga mansub bo'lganligi sabab joylarni nomlanishidagi farq u qadar katta emasdek tuyuladi. Toponimga aylangan so'z yangi ma'no kashf etish bilan ko'pincha tipik toponim shakl oladi, toponimlar uchun xos bo'lgan qo'shimchalar (suffiksalar) paydo bo'ladi. O'zbekiston toponimlari, xususan, oykonimlari uchun — iston (Sho'riston, Bog'iston), -iya (O'zbekiya, Mingiya), -kor (Paxtakor, Lalmikor), -cha (Quduqcha, Buloqchi) kabi affiksalar xarakterlidir. Hozirgi vaqtda geografik toponimlarda so'zlarning toponimga aylanmasdan oldingi ma'nosi aniq bilinib turadi, boshqa bir xil toponimlarda qisman saqlanib qolgan, uchinchi bir xil nomlarda bilinar bilinmas seziladi, yana bir xillarida esa umuman yo'qolib ketgan. Shuning uchun ham ma'nosi tushunarsiz bo'lgan toponimlarni har ikki tilda ham biroz o'zgartirib tushunarli qilish to'g'ri emas, chunki bunday nomlar qadimiy tillardan qolgan qimmatli lisoniy va tarixiy yodnomalar bo'lishi mumkin.

Adabiyotlar ro'yxati:

1. Yo'ldoshev I, Tunjay O'zturk. O'zbek va turk maqollari, iboralari. -Toshkent: 1998.
2. Sayfullava R, Mengliyev B, Qurbonova M. Hozirgi o'zbek tili, -Toshkent: Universitet, 2010.
3. Qurbonova M, Yo'ldoshev M. Matn tilshunosligi -Toshkent: Universitet, 2014.
4. O'zbek tilining izohli lug'ati, 1-2- jildlar, Москва: Русский язык, 1981;
5. Türkçe Sözlük, TDK, 9. Baskı, 1.-2. Ciltler, Ankara: 1998.
6. Büyük Rusça-Türkçe Sözlük, Москва, Русский язык, Multilingual, İstanbul: 1995.
7. Büyük Türkçe-Rusça Sözlük, Москва, 1977; İstanbul: 1995.
8. Aksoy Ö A. Atasözleri ve Deyimler Sözlüğü. 1. Atasözleri Sözlüğü, İstanbul: İnkilap Kitabevi. 1971,1974.
9. Aksoy, Ö. A., Bölge Ağızlarında Atasözleri ve Deyimler, 1. 2. Cilt, Ankara Üniversitesi Basımevi, Ankara:1996.
10. Albayrak N. Ansiklopedik Halk Edebiyatı Terimleri Sözlüğü. L.M.Yayınları, İstanbul: 2004.
11. Gür Alim, Ebüzziya Tefik. Hayat, Dil, Edebiyat, Basın, Yayın ve Matbaacılığa Katkıları, T.C. Kültür Bakanlığı Kültür Eserleri, Türk Tarih Kurumu Basımevi – Ankara: 1998.
12. Kartal N. Atasözleri ve Deyimler Sözlüğü, İnkilapYayınevi. -İstanbul: 1999
13. Türk Dünyası El Kitabı, Üçüncü Cilt.- Ankara:1992.
14. Yund, Kerim, Atasözlerinde Kişi Adları,Türk Yurdu, Cilt:6 Sayı:334, Ankara: 1967.
15. <http://wikipedia.com>
16. <http://ziyo.uz>
17. <http://school.com>
18. <http://referat.ru>
19. <https://tdk.gov.tr>
20. https://turkoloji.cu.edu.tr/pdf/ozbekce_bildiri.pdf

O'ZBEK TILIDAGI METAFORALARNING INGLIZ TILIGA TARJIMASIDA TARJIMA MUAMMOLARI

Saidova Go'zal Shuxratjon qizi,

Mirzo Ulug'bek nomidagi

O'zbekiston Milliy universiteti 3-kurs talabasi

E-mail: gozalsaidova51@gmail.com

Abdusattarova Durdona Alisher qizi

Mirzo Ulug'bek nomidagi

O'zbekiston Milliy universiteti 3-kurs talabasi

E-mail: abdusattarovadurdona84@gmail.com

Annotatsiya: Ushbu maqolada metaforalarni ingliz tiliga tarjima qilishda yuzaga keladigan lingvistik va madaniy tafovutlar, leksik-semantik noaniqliklar, ekvivalent topish qiyinchiliklari, idiomatik ifodalarning to'g'ri talqin qilinmasligi kabi muammolar batafsil ko'rib chiqilgan. Shuningdek, bu muammolarni bartaraf etish uchun kontekstual tafakkur, adaptatsiya, ekvivalentlik kabi yondashuvlardan foydalanish samarali natija berishi aniqlangan. Tadqiqot natijalari tarjimonlarga madaniyatlararo tafovutlarni hisobga olgan holda badiiy va emotsional mazmunni to'liq va aniq yetkazish yo'lida amaliy yordam beradi.

Kalit so'zlar: metafora, tarjima nazariyasi, lingvistik muammolar, madaniy tafovutlar, idiomatik ifoda, ekvivalentlik, semantik tafovut, adaptatsiya.

Abstract: This article thoroughly examines the challenges in translating metaphors from Uzbek into English, such as linguistic and cultural differences, lexical-semantic ambiguities, difficulties in finding equivalents and misinterpretation of idiomatic expressions. The study suggests the application of methods as contextual thinking, adaptation, and equivalence to overcome these issues. The results of the research provide practical guidance for translators accurately and fully conveying artistic and emotional content while considering intercultural differences.

Keywords: metaphor, translation theory, linguistic problems, cultural differences, idiomatic expression, equivalence, semantic difference, adaptation.

Аннотация: В данной статье подробно рассматриваются проблемы, возникающие при переводе метафор с узбекского языка на английский. В частности, обсуждаются лингвистические и культурные различия, лексико-семантические неопределённости, трудности в подборе эквивалентов и неправильная интерпретация идиоматических выражений. Также предлагаются методы решения этих проблем, такие как контекстуальное мышление, адаптация и эквивалентность. Результаты исследования оказывают практическую помощь переводчикам в точной и полной передаче художественного и эмоционального содержания с учётом межкультурных различий.

Ключевые слова: метафора, теория перевода, лингвистические проблемы, культурные различия, идиоматическое выражение, эквивалентность, семантическое различие, адаптация

Metafora hosila ma'no yuzaga kelishining eng faol shakllaridan biridir. Tilshunoslikda o'zaro o'xshashlik asosida, biror ma'noni boshqa bir ma'no orqali ifodalash natijasida ko'chimning metafora usuli yuzaga keladi. Bunda hosil qiluvchi va hosila ma'no referentlari o'rtasidagi semantik o'xshashliklar asosida yangi ma'no hosil qilinadi.

Metafora o'zbek tilida eng ko'p uchraydigan nom ko'chish usuli bo'lib, unda biror predmet, belgi, harakatning nomi boshqasiga o'zaro o'xshashligi asosida ko'chiriladi. Metaforaning asosiy ko'rinishlari:

Bir narsaning nomi boshqa bir narsaga shakliy o'xshashlik asosida ko'chiriladi: tish (odamning tishi) - tish (arraning tishi). Bir narsaning nomi boshqa bir narsaga o'rindagi o'xshashlik asosida ko'chiriladi: bosh (odamning boshi) - bosh (terakning boshi). Bir predmetga (ba'zan harakatga) xos belgining nomi boshqa bir predmetdagi belgiga ko'chiriladi: nordon (nordon olcha) - nordon (nordon gaplar). Bir predmetga xos harakatning nomi boshqa bir predmetning harakatiga ko'chiriladi: tozalamoq- (uyni tozalamoq) - tozalamoq (qalbni tozalamoq).

Imiy adabiyotlarda metaforaning bir necha xil tasniflari beriladi. Bu haqida Shohista Mamaraimovanning doktorlik ishida ham to'liq ma'lumotlar berib o'tilgan. M.Mirtojiyevning "O'zbek tili semasiologiyasi" nomli monografiyasida metaforalarning *oddiy* metafora, *personifikatsiya* va *sinesteziya* kabi turlari xususida keng ma'lumotlar berilgan. Bundan tashqari metaforaning *oddiy*, *jonlantiruvchi*, *sinestezik* va kognitiv kabi bir qator metaforalar turi mavjud. [M.Hakimova, 43]

O'zbekiston va boshqa mamlakatlar o'rtasidagi til va madaniyat farqlari, shuningdek, o'zbek va ingliz tillarining semantik xususiyatlari tufayli metaforalar tarjimasini maxsus e'tiborni talab qiladi. Ba'zan metaforalar boshqa tilga to'g'ri keladigan ekvivalentga ega bo'lmasligi mumkin, bu esa tarjimonlar uchun muayyan qiyinchiliklar tug'diradi. O'zbek tilidagi metaforalarning ingliz tiliga tarjimasini jarayonida yuzaga keladigan asosiy qiyinchiliklarni aniqlash va ularni bartaraf etish usullarini ishlab chiqish tarjima borasidagi dolzarb vazifalardan biridir. Bu esa, o'z navbatida, tarjimonlarga to'g'ri va madaniyatga mos tarjima qilishda yordam beradigan usul va yondashuvlarni taklif etadi.

Har bir til o'ziga xos grammatikasi, leksikasi va madaniy tizimiga ega ekanligi tarjima jarayonlarini qiyinlashtiradi. Shunday qilib, bir tilni boshqasiga tarjima qilishda yuzaga keladigan asosiy muammolarni aniqlash juda muhimdir. Ushbu muammolar asosan til tizimi, madaniy tafovutlar, leksik-semantik noaniqliklar, ekvivalent topish va idiomatik ifodalarning farqli talqin qilinishi bilan bog'liq. Bu muammolarni bartaraf etish har bir til semantikasini chuqur bilishni talab etadi.

Madaniy tafovutlar bilan bog'liq muammolar: o'zbek tilida metaforalar xalqning mentaliteti, urf-odatlar va tarixiy tajribasini aks ettiradi. Shuning uchun metaforalarni ingliz tiliga tarjima qilishda madaniy noaniqliklar yuzaga kelishi mumkin. Masalan, an'anaviy tandir pechining ochiq tomoni, ya'ni non yopiladigan yoki o'tin solinadigan joy – "tandirning og'zi". Agar "tandirning og'zi" to'g'ridan-to'g'ri tarjima qilinsa — "the mouth of the tandoor" bo'ladi. Biroq bu tarjima ingliz tilida o'zbek tilidagidek ma'no anglatmaydi, chunki ingliz tilida bunday madaniy tushuncha yo'q. Bunday paytda *Tandoor* (*tandir*) so'zi boshqa madaniyatlarda lakunani yuzaga keltiradi. Shuning uchun madaniy moslashtirish (cultural adaptation) zarur.

Leksik tafovutlar va semantik muvofiqsizliklar: har bir tilning lug'aviy boyligi o'ziga xos bo'lib, ayrim so'z yoki iboralar boshqa tilda to'g'ri ekvivalentga ega bo'lmasligi mumkin. Masalan, "aybini yuvmoq" birikmasida *yuvmoq* so'zi kimningdir qilgan xatosi yoki gunohini tuzatishga, o'zini oqlashga urinishi holati ma'nosida ifodalangan. Mazkur birikma so'zma-so'z tarzda ingliz tiliga "to wash one's fault" deb tarjima qilinishi mumkin, biroq ingliz tilida bu ifoda semantik jihatdan har doim ham ayni ma'noni anglatmaydi. Chunki ingliz tilida "wash" fe'li odatda jismoniy tozalashni bildiradi va bunday kontekstda ishlatilganda yuqoridagi ifodalangan ma'noni berolmaydi. Shu bois, bu birikmani kontekstdan kelib chiqqan holda "He tried to redeem himself" yoki "He atoned for his mistake" kabi ifodalar bilan tarjima qilish maqsadga muvofiqdir.

Ekvivalent topish bilan bog'liq muammolar: Metaforalarni tarjima qilish jarayonida yuzaga keladigan asosiy muammolardan biri bu — ekvivalent (muqobil) variant topish. O'zbek tilida ishlatiladigan "*Yuragi muzlab ketdi*" metafora usulida ma'no ko'chgan birlik odamning birdan cho'chib ketganini yoki dahshatli holatga tushganlik ma'nosini bildiradi. Uni ingliz tiliga

so‘zma-so‘z tarjima qilganda “His heart froze” ko‘rinishini oladi. Ingliz tilida bu ibora ayrim hollarda qo‘rquv yoki hayratni bildirishi mumkin bo‘lsa-da, u har doim ham o‘zbek tilidagi ma‘no yukini to‘liq ifodalay olmaydi. Shu bois tarjimada kontekstga qarab “He was terrified”, “He was stunned” kabi ifodalar qo‘llanishi mumkin.

Idiomatik va obrazli ifodalar bilan bog‘liq muammolari: Idioma (yoki frazeologik birlik) — bu doimiy tarkibga ega bo‘lgan, ko‘pincha ko‘chma ma‘noni anglatuvchi, barqaror ibora bo‘lib, uning umumiy ma‘nosi uni tashkil etgan leksik birliklarning alohida ma‘nolaridan kelib chiqmaydi. Idiomalar ko‘pincha obrazli ifoda bo‘lib, xalqning milliy tafakkuri, madaniy qadriyatlari va tarixiy an‘analarini o‘zida aks ettiradi. Idioma va obrazli ifodalar boshqa tilga so‘zma-so‘z tarjima qilinganda, asl ma‘no yo‘qolib ketishi yoki noto‘g‘ri talqin qilinishi mumkin. Masalan, “Ishtahasi qochmoq” birikmasida *qochmoq* so‘zi biror kishining emotsional holati yoki ruhiy tushkunligi sababli ovqat yeyishga ishtiyoqi yo‘qolganini bildiradi. Bu metaforani ingliz tiliga “He lost his appetite” deb tarjima qilish mumkin, biroq bu tarjima asos tildagi to‘liq konnotativ ma‘noni ochib bera olmasligi mumkin. Ayrim hollarda bu birikma emotsional tushkunlik g‘am yoki ruhiy charchoq holatini ham ifodalashi mumkin. Shu sababli, tarjimada “He was too upset to eat”, “He had no desire to eat due to stress” yoki “His grief took away his appetite” kabi kontekstga moslashtirilgan ifodalarning ishlatilishi tog‘ri tarjimini yuzaga chiqaradi.

Tarjima jarayoni nafaqat tilning grammatik va leksik jihatlarini, balki madaniyatlararo tavofutlarni ham hisobga olishni talab etadi. Tarjimonning asosiy vazifasi faqat so‘zlarni tarjima qilishda emas, balki maqsadli til va madaniyatdagi ma‘noni to‘liq va aniq yetkazishda ham ko‘rinadi. Har bir tarjimon o‘z ishini mukammal bajarishga intilayotganda, leksik, grammatik va madaniy farqlarni e‘tiborga olishi lozim. Chunki har bir tilning o‘ziga xos struktura va semantik xususiyatlari bor, shu bilan birga, madaniy kontekst, iboralarning ma‘nosi, odatlar, an‘analar va o‘zgacha dunyoqarashlar tarjima jarayonida muhim rol o‘ynaydi. Shuningdek, tarjimonlar o‘zlarining ijodiy yondashuvlarini ham qo‘llashlari zarur. Ijodiy yondashuv nafaqat tarjimaning to‘g‘ri va aniq bo‘lishini ta‘minlash, balki matnning o‘ziga xosligini va originalining ruhini saqlab qolish uchun ham muhimdir. Madaniy farqlarni yengish uchun zamonaviy usullarni ishlab chiqish va ulardan samarali foydalanish tarjimonlarning professionallik darajasini oshiradi. Masalan, tarjimadagi badiiy va idiomatik ifodalarni muvofiqlashtirishda, tarjimonlar original matndagi hissiy va badiiy elementlarni to‘liq saqlashga harakat qilishi kerak, shunda u maqsadli tilga mos tarzda yetkaziladi. Tarjima jarayonida yuzaga keladigan muammolarni hal qilish uchun nafaqat til bilimlari, balki madaniyatshunoslik va kontekstual tafakkur ham talab etiladi. Tarjimonlar har doim maqsadli til va madaniyatning sharoitlariga mos ravishda talqin qilish va o‘zgarishlar kiritish orqali, asl matnning ma‘nosini saqlab qolishlari zarur. Misol uchun, madaniyatlararo tafovutlar, an‘analar yoki biror tushuncha boshqa tilda boshqacha talqin qilinishi mumkin, shuning uchun tarjimonlar turli yondashuvlarni qo‘llashlari kerak, masalan, adaptatsiya, ekvivalentlik kabi metodlarni. Bundan tashqari, tarjima jarayonida yuzaga keladigan semantik, pragmatik va badiiy muammolarni hal qilish uchun kontekstual tafakkur ham katta ahamiyatga ega. Tarjimon faqat tilning grammatik va leksik qoidalari bilan cheklanmasdan, tilning madaniy va ijtimoiy sharoitlarini ham hisobga olib, maqsadli tilga eng mos va aniq tarjimini taqdim etishi kerak. Shunday qilib, tarjimonlarning asosiy vazifasi nafaqat so‘zlarni to‘g‘ri tarjima qilish, balki har bir elementni madaniy kontekstda o‘rganib, tarjima jarayonini yanada boyitishdan iboratdir.

Zamonaviy tarjima jarayonlarida kontekstual tafakkur va madaniyatshunoslik bilimlarini qo‘llash, tarjimaning muvaffaqiyatini ta‘minlashda muhim omil hisoblanadi. Shunday qilib, tarjimonlarning asosiy vazifasi – tilning nozik jihatlarini va madaniy tafovutlarni hisobga olib, yuqori sifatli tarjima taqdim etishdir.

Foydalanilgan adabiyotlar:

1. Abdurakhmonova, N. & Shamieva, N. (2024, November). Creating an English-Uzbek Bilingual Thesaurus of Frequently Used Adjectives in Uzbek Corpus. In 2024 IEEE 3rd

- International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE) (pp. 1640-1644). IEEE.
2. Abduraxmonova, N. & Vasliddinova, K. (2024). O‘zbek tili parallel korpusining lingvistik masalalari. Computer linguistics: problems, solutions, prospects,1(1).
 3. Baker, M. (2018). In other word: A coursebook on translation. Routledge.
 4. Hakimova. M. (2022). O‘zbek tilida “holat” ma’noli abstrakt nomlar. Uzbekistan: Language and Culture, 2(2)
 5. Ismailov, A. S. Shamsiyeva, G. Abdurakhmonova, N., & Navoi, A. (2021). Statistical machine translation proposal for Uzbek to English. Science and Education, 2(12), 212-219
 6. Lakoff, G. & Johnson, M. (2008). Metaphors we live by. University of Chicago press.
 7. Mirtojdiyev M. M. “O‘zbek tili semasiologiyasi” Toshkent (2017).
 8. Newmark, P. (1988). A textbook of translation (Vol. 66, pp. 1-312). New York: Prentice hall.

XORAZM SHEVALARI SINONOMIK LUG'ATINING MA'LUMOTLAR BAZASINI ISHLAB CHIQISH

Sharipov Maqsud Siddiqovich

Texnika fanlari nomzodi, dotsent

Abu Rayhon Beruniy nomidagi Urganch Davlat Universiteti

E-mail: maqsbek72@gmail.com

Kurbanova Lola Ulug'bek qizi

Teacher, Urganch State University

E-mail: lolakompyuterlingvistikasi@gmail.com

Atanazarova Madina

Urganch RANCH Texnologiya Universiteti magistranti

E-mail: atanazarovamadina677@gmail.com

Annotatsiya: Shevalar bilan bog'liq bo'lgan leksik, fonetik va semantik tafovutlarni raqamli tizimlarga kiritish — tilshunoslikda yangi yo'nalishlar va tadqiqotlar uchun keng imkoniyatlar yaratmoqda. Ushbu maqolada xorazm shevalariga oid sinonim so'zlarni elektron tarzda tizimlashtirish va ular asosida zamonaviy ma'lumotlar bazasini yaratish jarayoni yoritiladi. Lug'at loyihasi uchun dastlab adabiy o'zbek tilidagi so'zlar va ularning xorazmcha sinonim shakllari yig'ib chiqildi. Keyinchalik bu ma'lumotlar MySQL yordamida ma'lumotlar bazasida saqlanishi uchun strukturaviy jadval holiga keltirildi. Yaratilgan Excel va SQL fayllar muallifning GitHub hisobiga yuklash maqsad qilingan bo'lib, ushbu maqoladagi yondashuvdan foydalanmoqchi bo'lgan tadqiqotchilar uchun ochiq manba sifatida taqdim etiladi.

Kalit so'zlar. O'zbek tili, Xorazm shevasi, ma'lumotlar bazasi, lingvistik manba, MySQL.

Annotation: Incorporating lexical, phonetic, and semantic differences related to dialects into digital systems creates vast opportunities for new directions and research in linguistics. This article highlights the process of electronically systematizing synonymic words from the Khorezm dialects and developing a modern database based on them. For the dictionary project, standard Uzbek words and their Khorezm dialectal synonyms were first collected. These data were then structured into tabular form for storage in a MySQL database. The generated Excel and SQL files are intended to be uploaded to the author's GitHub account and made available as open source for researchers who wish to use the approach in this article.

Key words. Uzbek language, Khorezm dialect, database, linguistic resource, MySQL.

Аннотация: Лексические, фонетические и семантические различия, связанные с диалектами, открывают широкие возможности для новых направлений и исследований в лингвистике при их внедрении в цифровые системы. В данной статье освещается процесс систематизации синонимичных слов, относящихся к Хорезмским диалектам, в электронном виде и создание на их основе современной базы данных. Для проекта словаря сначала были собраны слова на литературном узбекском языке и их хорезмские синонимы. Затем эти данные были структурированы в виде таблиц для хранения в базе данных с использованием MySQL. Сгенерированные файлы Excel и SQL предназначены для загрузки в учетную запись автора на GitHub и предоставления в качестве открытого исходного кода исследователям, желающим использовать подход, описанный в этой статье.

Ключевые слова. Узбекский язык, хорезмский диалект, база данных, лингвистический ресурс, MySQL.

So‘nggi yillarda O‘zbekistonda kompyuter lingvistikasi sohasida sezilarli yutuqlarga erishilmoqda. Bu yo‘nalish, tilni raqamli texnologiyalar orqali o‘rganish, tahlil qilish va amaliyotga joriy etishni maqsad qiladi. Xususan, sun'iy intellekt asosidagi suhbat botlari, ovoqli yordamchilar va boshqa ko'plab qurilmalardagi dasturiy ta'minotlar foydalanuvchilarga og‘zaki nutq orqali muloqot qilish imkonini bermoqda. Bunday tizimlarning samarali ishlashi uchun esa tilning turli qatlamlarini, jumladan, shevalarni chuqur o‘rganish va ularning lingvistik bazalarini yaratish muhim ahamiyatga ega. Shevalar, ayniqsa, tilning og‘zaki shaklida keng qo‘llaniladi va ular mintaqaviy xususiyatlarni aks ettiradi. Shu sababli, kompyuter lingvistikasi sohasida shevalarni o‘rganish, ularning leksik, fonetik va grammatik xususiyatlarini aniqlash, sun'iy intellekt tizimlarining tilni to‘g‘ri tushunish va qayta ishlash qobiliyatini oshiradi.

Xorazm shevasi o‘zbek tilining muhim shevalaridan biri bo‘lib, u asosan Xorazm viloyati va unga yaqin hududlarda keng tarqalgan. Bu sheva turkiy tillar oilasining o‘g‘uz guruhiga mansub bo‘lib, o‘zining fonetik, leksik va grammatik xususiyatlari bilan ajralib turadi. Masalan, so‘z boshida "k" tovushining "g" tovushiga aylanishi (masalan, "kel" – "gel"), "t" tovushining "d" tovushiga aylanishi (masalan, "til" – "dil") kabi fonetik o‘zgarishlar kuzatiladi. Xorazm shevasi o‘zining boy leksikasi, qadimiy so‘zlar va iboralarning saqlanib qolganligi bilan ham ajralib turadi. Bu sheva o‘zbek tilining tarixiy taraqqiyotini o‘rganishda muhim manba hisoblanadi.

O‘zbek tilining shevalari, jumladan, Xorazm shevasi bo‘yicha bir qator ilmiy tadqiqotlar olib borilgan. Mahmud Koshg‘ariy "Devonu lug‘otit turk" [1] asarida turkiy tillarning hududiy-geografik jihatdan yashash manzillari va lingvistik xususiyatlarini alohida e‘tiborga olib, turkiy tillarni ikkita guruhga bo‘ladi: 1) turk yoki xoqoniy turkchasi tillari; 2) o‘g‘uz tillari. O‘zbek tilining o‘g‘uz lahjasi va uning ayrim fonetik xususiyatlari haqida olib borilgan tadqiqotlarda Xorazm shevalarining fonetik, leksik va grammatik xususiyatlari chuqur o‘rganilgan.

Ushbu maqolada [2] Xorazmdagi o‘g‘uz shevalarining gastronomik leksikasi tahlil qilingan. Bu maqolada [3] turkiy tillar va Xorazm shevalari o‘rtasidagi bog‘lanishlarni aniqlash maqsadida qiyosiy-tarixiy va etimologik lug‘at tuzilgan bo‘lib, u NLP va qiyosiy tadqiqot usullari yordamida tillarning tarixiy ildizlarini ochib beradi. Ushbu maqolada [4] Xorazm shevasidan O‘zbek adabiy tiliga avtomatik tarjima qilish uchun maxsus ma‘lumotlar bazasi yaratilgani va uning metodologiyasi yoritiladi. Tadqiqot shevaning fonetik, morfologik va sintaktik xususiyatlarini hisobga olib, tarjima modellari samaradorligini oshirishga qaratilgan. Xorazm shevasi bilan turk tili va o‘zbek tilining boshqa shevalari orasidagi leksik o‘xshashliklarni ko‘rsatib, aynan ismlar va fe‘llarga oid misollar asosida bu shevaning turk tiliga yaqinligini ta‘kidlangan [5]. Matnlarda ot va olmosh so‘z turkumlarini aniqlash uchun ularning tup shakldagi bazasi va ularning qo‘shimchalar bazasi shakllantirilgan [6]. Stemmingdan foydalangan holda matnlarni belgilash jarayoni amalga oshirilgan, turli sohalarga tegishli bo‘lgan 23 000 atrofida so‘zlar testlab ko‘rilgan va 89.78% aniqlikka erishilgan [7].

Ushbu loyiha doirasida Xorazm shevasiga oid sinonimik lug‘atni yaratish uchun so‘zlar turli manbalardan yig‘ildi. Avvalo, adabiy o‘zbek tilidagi asosiy so‘zlar turli ilmiy lug‘atlar, darsliklar va nasriy asarlardan tanlab olindi. Bu so‘zlar adabiy norma sifatida xizmat qiladi.

Xorazmcha sinonimlar esa, asosan, ijtimoiy tarmoqlar (Telegram, Facebook, Instagram) orqali joylashtirilgan postlar, izohlar, muloqotlar tahlili orqali yig‘ib olindi. Bu usul shevada faol ishlatilayotgan hozirgi zamonaviy og‘zaki nutq namunalarini jamlashga yordam berdi. Shuningdek, intervyular va muloqotli suhbatlar asosida yozib olingan ma‘lumotlar ham bu lug‘atni boyitishda xizmat qildi.

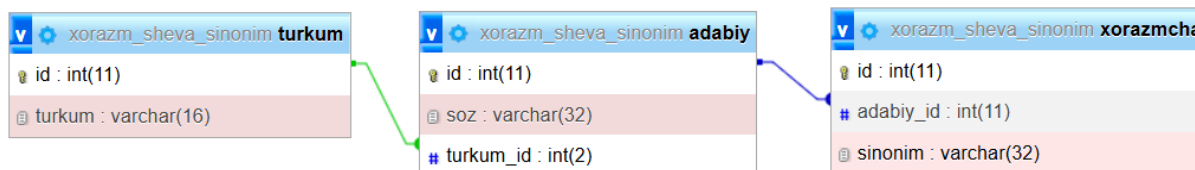
Yig‘ilgan so‘zlar 1-jadvaldagi kabi yozib chiqildi. Har bir satrda adabiy so‘z, uning so‘z turkumi va unga mos keluvchi Xorazm shevasidagi 3–5 sinonimi berilgan.

№	Adabiycha so‘z	So‘z turkumi	Xorazmcha				
			s1	s2	s3	s4	...
1	yuz	ot	bet	aft	bashara	chehra	
2	tayoq	ot	toyoq	cho‘p	chyuq		

3	go'sht	ot	go'sh	et			
4	yashirin	sifat	im-sim	bildirmin	indamin	yoshinib	
5	mana bu	olmosh	ibi	ishi	inobi	onashi	
6	yanglishmoq	fe'l	olvuramoq	obtiramog	odoshmoq		
7	tez	ravish	darrov	tez-tez			
8	to'qqiz	son	do'qqiz				

1-jadval. Xorazmcha so'zlarining sinonimik lug'ati.

Yig'ilgan ma'lumotlarni tizimlashtirish uchun MySQL muhitida ma'lumotlar bazasi yaratildi. Quyida bu bazaning diagramma shaklidagi modeli keltirilgan:



1-chizma. Xorazmcha so'zlarning sinonimik so'zlarning ma'lumotlar bazasi.

Ma'lumotlar tuzilmasi quyidagi jadvalardan iborat:

- adabiy – adabiycha so'zlar jadvali
- xorazmcha – shevadagi sinonimlar jadvali
- turkum – so'z turkumlari

Bunday struktura tanlanishining asosiy sababi – ma'lumotlarni modullashtirish, kengaytirish imkoniyati, va keyinchalik qidiruv tizimlarini optimallashtirish imkoniyatining mavjudligidir.

Xorazm shevasi o'zining boy leksikasi, fonetik va grammatik xususiyatlari bilan o'zbek tilining muhim shevalaridan biri hisoblanadi. Kompyuter lingvistikasi va sun'iy intellekt sohalarida bu shevaning lingvistik bazasini yaratish, tilni chuqurroq o'rganish va texnologik tizimlarda qo'llash imkonini beradi. Shu sababli, Xorazm shevasini ilmiy jihatdan o'rganish va uning ma'lumotlar bazasini yaratish dolzarb masala hisoblanadi.

Foydalanilgan adabiyotlar.

1. Wikipedia. https://uz.wikipedia.org/wiki/Devonu_lug'otit_turk.
2. Yo'ldoshevna BS. Xorazm viloyatidagi O'G 'UZ shevasi gastronomik leksikasining lisoniy xususiyatlari. tanqidiy nazar, tahliliy tafakkur va innovatsion g 'oyalar 2025; 1: 101–103.
3. Rahimbayeva M, Rajapboyeva Y. turk tili va xorazm shevalari o 'xshash leksemalarining qiyosiy-tarixiy va etimologik lug 'atini elektron shaklda yaratish. COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS 2024;
4. Siddiqovich S.M., Kizi KLU. Developing a database for translating texts from the Khorezm dialect into uzbek literary language. Electronic journal of actual problems of modern science, Education and Training: 23
5. Saburov b. Turk tili va Xorazm shevasidagi isim so 'z turkumiga oid o 'zaro ma'nodosh birliklar. International scientific journal of Biruni 2023; 2: 76–85
6. Sharipov M, Mavlonova M, Sobirov O. O'zbek tilidagi matnlarda ot va olmosh so'z turkumlariga tegishli so'zlarni aniqlovchi algoritm yaratish. Computer Linguistics: Problems, Solutions, Prospects 2023; 1.
7. Sharipov M, Kuriyozov E, Yuldashev O, Sobirov O. UzbekTagger: The rule-based POS tagger for Uzbek language. arXiv preprint arXiv:230112711 2023;

JAHON TILSHUNOSLIGIDA PRAGMALINGVISTIKA VA PRAGMATIK TAHLIL

Tursunova Sojidxon Axmadjon qizi

Alisher Navoiy nomidagi Toshkent

davlat o'zbek tili va adabiyoti universiteti tayanch doktoranti.

E-mail: sojidxatursunova0525@gmail.com

Annotatsiya: ushbu maqolada pragmatika va uning jahon tilshunosligidagi o'rni, dunyoga kelish tarixi, predmeti, obyekti, ushbu soha o'rganadigan masalalar; pragmatika terminining ilmiy muomalaga kiritilishi, pragmatik tahlil, pragmatik tahlilda ironiyaning o'rni va boshqa bir qator sohaga oid ma'lumotlar keltirilgan. Maqola Ch.Pirs, U.Djems, D.Dyun, Ch.Morris J.R.Ostin, J.R.Serl, Z.Vendler, V.V.Petrov, N.D.Arutyunova, Y.V.Paducheva kabi tilshunoslarning pragmalingsvistikaga oid fikrlari bilan boyitilgan.

Kalit so'zlar: pragmatika, pragmalingsvistik tahlil, pragmatikaning predmeti, nutqiy akt nazariyasi, ironiya va boshqalar.

Аннотация: В данной статье рассматриваются прагматика и её роль в мировой лингвистике, история её возникновения, предмет и объект изучения, а также основные вопросы, исследуемые в данной области. Освещается введение термина «прагматика» в научный оборот, сущность прагматического анализа и роль иронии в таком анализе, наряду с другими важными аспектами. Статья обогащена взглядами таких известных лингвистов, как Ч. Пирс, У. Джемс, Д. Дьюи, Ч. Моррис, Дж. Р. Остин, Дж. Р. Серль, З. Вендлер, В.В. Петров, Н. Д. Арутюнова, Ю. В. Падучева.

Ключевые слова: прагматика, прагмалингвистический анализ, предмет прагматики, теория речевых актов, ирония и другие.

Annotation: In this article, data is provided about pragmatics and its role in global linguistics, the history of its emergence, its subject and object, the tasks studied in this field, the introduction of the term "pragmatics" into scientific discourse, pragmatic analysis, the role of irony in pragmatic analysis, and other related topics. The article is filled with opinions, which are related to pragmalinguistics, of linguists like Ch.Pris, U.Djems, D.Dyun, Ch.Morris, J.R.Ostin, J.R.Serl, Z.Vendler, V.V.Petrov, N.D.Arutyunova, Y.V.Paducheva.

Keywords: pragmatics, pragmalinguistic analysis, the subject of pragmatics, theory of speech action, irony and others.

Jahon tilshunoslik ilmida yangi ilmiy paradigmalarning shakllanishi inson tafakkuri va nutqiy faoliyati bilan bog'liq masalalarni o'rganish zaruriyatini belgilab berdi. Ijtimoiy hayotda kishilar o'rtasidagi o'zaro aloqa jarayoni – kommunikatsiyani yuzaga chiqaruvchi nutq birliklarining pragmatik tadqiqiga e'tibor kuchaydi. Tobora o'sib, rivojlanib borayotgan fan sohada yangilik sifatida til birliklarining nutqiy voqelanishi ularning nutqiy vaziyat, kontekst bilan bog'liq muammolarini o'rganishga kirishdi. Bu esa fanning antroposentrik mohiyat kasb etishiga asos bo'ldi hamda tilshunoslikning psixolingvistik, sotsiolingvistik, pragmalingsvistika, kognitiv tilshunoslik, lingvokulturologiya kabi inson omili bilan bog'liq bo'lgan yangi sohalarini shakllandi. Biz mana shu yangi sohalar qatoridan munosib o'rin egallagan pragmalingsvistika haqida so'z yuritmoqchimiz.

Pragmatika tilshunoslik fanining yangi bir nazariy va amaliy tarmog'idir. U insonning ijtimoiy faoliyatini o'zida aks ettiruvchi nutqiy jarayon, nutq ishtirokchilariga xos kommunikativ niyat, nutqiy vaziyatning ta'siri bilan aloqador masalalarni o'rganadi. "Pragmatikaning manbasi Ch.Pirs, U.Djems, D.Dyun, Ch.Morris kabi taniqli faylasuf olimlar nomi bilan bog'liqdir. Chunki ularning ishlarida (XIX asrning oxiri XX asrning boshlarida) belgilar sistemasi va lingvistik belgi funktsionalligi xususidagi g'oyalar o'rtaga tashlanib, semiotika haqidagi asosiy tushunchalar

aniqlanadi, sintaktika, semantika va pragmatika o‘rtasidagi o‘zaro farqlar ko‘rsatiladi”.[Hakimov M., 2013:6] .

Pragmatika yunoncha pragmos so‘zidan olingan bo‘lib, “ish”, “harakat” degan ma’nolarni bildiradi. Tilshunoslikda “pragmatika” terminining paydo bo‘lishi falsafaning pragmatizm yo‘nalishi bilan aloqador sanaladi. Pragmatika termini va pragmatik g‘oyalarning aynan tilshunoslik sohasida shakllanishiga katta hissa qo‘shgan olim avstriyalik psixolog Karl Lyudvig byullerdir.

Pragmatika tilshunoslikda dastlab mustaqil fan shaklida emas, sohaning ma’lum bir qismi sifatida o‘rganilgan. Tilshunoslar pragmalingvistikaga turli nazariy qarashlar bilan yondashgan. Pragmatikaning “dunyoga kelishi” falsafa bilan bog‘liq bo‘lsa-da, ammo pragmatizm yo‘nalishining asoschilaridan biri Ch.Pirs ushbu falsafiy tizimning asosiy g‘oyasi semiotik belgining (lisoniy belgi ham nazarda tutilgan) ma’no-mazmunini ushbu belgi vositasida bajarilayotgan harakatning samarasi, natijalari, muvaffaqiyati bilan bog‘liq holda o‘rganilishi va bu tamoyil belgi nazariyasi doirasida kommunikativ faoliyat subyekti omilini inobatga olish lozimligini uqtirdi.

Ch.Pirsning g‘oyalari falsafiy pragmatizm va lingvopragmatika sohalari uchun nazariy asos bo‘la oldi. Tilga belgilar asosida yondashish lisoniy birliklarni nutqiy voqelanishdan “uzib qo‘ygan”dek tuyulsa-da, ammo Pirs nazarda tutgan semiotik belgilar pragmatikani lingvistikaga yaqinlashtirdi. [Xujamkulova M., 2023:10].

Pragmatika real muloqot jarayonlarida belgilarning xatti-harakatlarini o‘rganadi. Ch.Morris ritorikani “pragmatikaning asosiy salafi” deb ataydi. “Ko‘pgina belgilarning (ehtimol hammasining) interpretatorlari tirik organizmlar bo‘lgani uchun, – deb yozadi Ch.Morris, pragmatikaning asosiy xususiyati semiozisning biotik jihatlarini, boshqacha qilib aytganda, belgilarning ishlashi paytida kuzatiladigan barcha psixologik, biologik va ijtimoiy hodisalar bilan bog‘liq”. [Madaminova M., 2023:5].

Pragmatika termini XX asrning 30-yillarida Ch.Morris tomonidan ilmiy muomalaga kiritilgan. U semiotikani quyidagi tarkibiy qismlarga ajratadi: semantika, sintaktika va pragmatika.

J.R.Ostin, J.R.Serl va Z.Vendlerlarning nutqiy akt nazariyasi ta’limoti ta’sirida lingvistik pragmatikaning shakllanishi uchun ma’lum asoslar yaratildi. Nutqiy aktlar yig‘indisi nutqning mazmunini tashkil qiladi. So‘zlovchining munosabati esa nutqiy aktlar vositasida shakllanadi. Bunda inson tilining ibtidoiy shakllanishidan tortib to uning murakkab mazmuniy tuzilishiga qadar bo‘lgan lingvistik xususiyatlari nutqiy akt muammolari sifatida talqin etiladi. Nutqiy akt nazariyasiga xos har bir tadqiqiy talqinda nutq subyektiining olamni bilish va his qilishdagi tasavvur ko‘rsatkichlarining nutqdagi darajalari hisobga olinadi. Nutqiy akt nazariyasiga xos mana shunday o‘zak tadqiqiy talqin lingvistik pragmatika tarmog‘ining shakllanishi uchun asosiy mezon vazifasini bajaradi. [Hakimov M., 2013:6].

So‘zlovchining nutqi bamisoli bir ko‘zgudir. Uning nutqi orqali so‘zlovchining dunyoqarashi, fikrlashi, ijtimoiy holati, yashash tarzi, jamiyatdagi o‘rni va yana boshqa bir qator ma’lumotlarni bilib olish mumkin. Nutq kishilarning kim ekanligidan hamda asl maqsadi nima ekanligidan darak berib turadi.

Pragmatika – bu til tashuvchisi va til o‘rtasidagi munosabatni o‘rganadigan fan. Pragmatika ikki xil muammoni tahlil qiladi: birinchidan, nutqiy akt va nutq “mahsulotlari”ning qiziqarli turlarini aniqlash; ikkinchidan, nutqiy kontekstda, muayyan jumlada qanday propozitsiya ifodalanishini aniqlashga ta’sir qiluvchi belgi va xususiyatlarni tavsiflash.

Tasdiq, buyruq, kontrafakt so‘zlashuv, talab, taxmin va rad etish, iltimos, e’tiroz, bashorat, va’da, chaqiriq, mulohaza, tushuntirish, haqorat, xulosa, taxmin, umumlashtirish, javob berish, aldash – bularning barchasi nutqiy akt turlaridir. Pragmatik tahlilning vazifasi nutqiy aktning muvaffaqiyatli (ba’zan normal) amalga oshishi uchun zaruriy va yetarli vaziyatni aniqlashdir. [Madaminova M., 2023:7].

Pragmatik tahlilda ironiyaning o‘rni ham beqiyosdir. So‘zlarimizning isbotini quyidagi misolda ham ko‘rishimiz mumkin.

- “Jo‘raqulboyning inisi To‘raqulboy kattaroq o‘qishga kirarmish”, deb eshitib edik. Qaysi kuni cho‘lda ko‘rib qoldim... cho‘ldagi o‘qishda, shekilli?
- Toshkentda emas ekanmi? Kari-fari olamga sig‘mas edi-ku uning, nima bo‘pti?
- Yiqilibdi-da.
- U yoq-bu yog‘i sinmabdimi ishqilib?
- Har qalay, biroz lat yegan bo‘lsa kerag-ov... [Erkin A‘zam, 2007:168].

Erkin A‘zamning “Fayzining amakisi” asaridan olingan ushbu parchada To‘raqulboyning ortiqcha maqtanishi uning mazax qilinishiga sabab bo‘lmoqda. Kattaroq o‘qishga kiraman deb cho‘ldagi o‘qishda “o‘qiyotgan” qahramonimiz “yiqilib” hammaga kulgiga qoldi. “Lat yegan bo‘lsa kerag-ov”, deya piching qilishlari ham shundan. Keltirilgan misolda ham masxaralash, mazax ochiq-oydin ko‘rinib turibdi, shundan kelib chiqsak, keltirilgan namunani ilmoq sifatida ham olishimiz mumkin. Yuqoridagi fikrlar zahirida “Behuda karillab, baland dorga osilishning, o‘ziga ortiqcha baho berishning oqibati shunday bo‘ladi”, degan pragmatik mazmun yashiringan.

Lingvistik pragmatika nazariyasi haqidagi ilk ma‘lumotlar tilshunoslik fanining yangi tarmog‘i sifatida o‘tgan asrning 60-70-yillarida boshlangan bo‘lsa ham, bu yo‘nalishning paydo bo‘lishiga XIX asrning oxiri hamda XX asrning boshlarida paydo bo‘lgan g‘oyalar turtki bo‘lgan.

60-70-yillarda til belgilarining nutqdagi amaliy qo‘llanishiga xos kontekstual xususiyatlarini o‘rganish bilan tilshunoslikda pragmatik yo‘nalish yuzaga kela boshladi. Lingvistik pragmatikaning shakllanishi yirik olim L.Vitgenshteynning falsafiy qarashlari bilan bog‘liq bo‘lib, aynan uning ishlarida pragmatika nazariy semiotik tadqiqotning mustaqil sohasi sifatida ko‘rsatib o‘tiladi. [Hakimov M., 2013:7].

1970-yilda Dordrext shahrida “Tabiiy tillar pragmatikasi” (Pragmatics of natural languages) mavzusida xalqaro konferensiya bo‘lib o‘tdi. Ushbu anjumanda o‘qilgan ma‘ruzalar to‘plamining muharriri, Tell-Aviv universitetining professori M.Bar-Hillelning “So‘z boshi”da qayd qilishicha, anjuman ishtirokchilari yakdillik bilan “Tabiiy til vositasida bajariladigan muloqotning pragmatik xususiyatlari ushbu muloqotning sintaktik va semantik xususiyatlari kabi lingvistik nazariya doirasida o‘rganilishi lozim”, degan xulosaga kelishdi. [Safarov Sh., 2018:56].

Pragmatikaning tilshunoslik fanining yangi tarmog‘i sifatida shakllanib, tan olinishi 70-yillarning ikkinchi yarmiga to‘g‘ri keladi.

1977-yildan Amsterdamda lingvistik pragmatikaga bag‘ishlangan maxsus jurnal nashr qilina boshlandi. Jurnal “Nutqshunoslik” va “Tekstologiya”ning turli yo‘nalishlari bilan bog‘liq masalalarni yoritishga bag‘ishlangan maqolalarni to‘plam sifatida e‘lon qilishni o‘z oldiga maqsad qilib qo‘ydi. [Hakimov M., 2013:7].

Tilshunos olim V.V.Petrovning til haqidagi quyidagi tezisi lingvopragmatikaning boshlang‘ich tadqiqiy chegarasini aniqlash uchun lingvistik asos bo‘ladi: “Inson qaysidir tilga xos semantik bilimlarsiz so‘zlashi yoki anglashi mumkin emas, ammo o‘sha tilga xos semantik bilimlar egasi bo‘lishning o‘zi ushbu tilda so‘zlash va anglash uchun kifoya qiladi degani emas, masalan, musiqali pyesa mazmunidan ogohlik uning ijrosi uchun yetarli bo‘lmaydi, buning uchun, eng asosiysi, ijro mahorati, layoqati (“умение играть”)ga ham ega bo‘lish lozim”.

Lingvistik pragmatikaning shakllanishi xususidagi asosiy ma‘lumotlar N.D.Arutyunova, Y.V.Paduchevalarning ishlarida beriladi. Ular an‘anaviy tilshunoslikka oid juda ko‘p ishlarda semantika ensiklopedik bilimlarga tenglashtirilgani, pragmatika to‘g‘risida esa umuman tasavvurning yo‘qligini ta‘kidlaydilar. Keyinchalik esa semantika masalalarini tadqiq etish jarayoni pragmatik hodisalarga bo‘lgan qiziqishni kuchaytirdi. [Hakimov M., 2013:8-9].

Xulosa qilib aytishimiz mumkinki, pragmalingvistika jahon tilshunosligida o‘z o‘rniga ega sohalardan biridir. U ko‘plab tilshunoslarning qiziqishlariga sabab bo‘lgan. Qolaversa, tilshunoslarning izlanishlari bois ushbu soha qimmatli ma‘lumotlar bilan boyib borgan.

Foydalanilgan adabiyotlar:

1. Erkin A‘zam. Ertak bilan xayrlashuv. Toshkent: O‘zbekiston milliy ensiklopediyasi. Davlat ilmiy nashriyoti, 2007. –B. 199.

2. Madaminova M. Matn pragmatik tavsifida sintaktik birliklarning o‘rni. – Toshkent, 2023. – B.99.
3. Xujamkulova M. Sodda gapda kommunikatsiya va pragmatika munosabati. – Toshkent, 2023. – B.140.
4. Сафаров Ш. Прагмалингвистика. – Тошкент, 2018. – В.320.
5. Ҳақимов М. Ўзбек прагмалингвистикаси асослари. – Тошкент: Академнашр, 2013. – В.172.

DISKURS TUSHUNCHASI VA UNING INTERTEKSTUALLIK HODISASI BILAN BOG'LANISHI

Ibodullayeva Kunduz

O'zMU Kompyuter lingvistikasi va
amaliy tilshunoslik kafedrası o'qituvchisi

Annotatsiya: Diskurs tushunchasi tilshunoslikda keng iste'molda bo'lgan konsepsiyalar qatoriga kirib ulgurdi. Adabiyot, ijtimoiy-maishiy, rasmiy, ilmiy sohalarda uchraydigan diskursning qo'llanish doirasi, lingvopragmatik, kognitiv xususiyatlarini tahlil qilish yangi yondashuvlarni va qoidalarini talab qiladi. Ushbu tadqiqotda diskurs tushunchasi va uning qamrovini sharhlash, mohiyatini va funksiyalarini ochib berish va diskursning intertekstuallik hodisasi bilan bog'liqligi haqida mulohazalar bildirilgan.

Kalit so'zlar: Diskurs tahlili, ilmiy diskurs, kognitiv tilshunoslik, antroposentrik paradigma, kommunikativ-pragmatik yondashuv, ilmiy diskursda intertekstuallik

Annotatsion: The concept of discourse has become one of the widely used concepts in linguistics. The analysis of the scope of application, linguopragmatic, and cognitive features of discourse encountered in literature, social-domestic, official, and scientific fields requires new approaches and rules. In this study, comments are made on the concept of discourse and its scope, revealing its essence and functions, and the connection of discourse with the phenomenon of intertextuality.

Keywords: Discourse analysis, scientific discourse, cognitive linguistics, anthropocentric paradigm, communicative-pragmatic approach, intertextuality in scientific discourse

Dunyo tilshunosligida nutqiy tuzilmalar, diskursning tilda shakllanish va voqelanishi, umumiy tavsifi hamda xususiy ko'rinishlariga doir ilmiy izlanishlar ko'plab tadqiqotchilarni o'ziga jalb qilib kelgan. Hozirgi kunga qadar diskurs masalasining turli ko'rinishlari (siyosiy, ijtimoiy, ta'limiy, diniy, ilmiy, badiiy, ommaviy) o'rganilmoqda. Ushbu tadqiqotlarda diskurs tushunchasi va uning qamrovini sharhlashga, mohiyatini va funksiyalarini ochib berishga harakat qilinganligi va bu boradagi munosabatlarning xilma-xilligini kuzatish mumkin. Shu ma'noda, nutqni shakllantiruvchi birliklar va ularni diskursiv jihatdan tadqiq qilishga e'tibor kuchaymoqda. Xususan, diskurs tushunchasi tilshunoslikda keng iste'molda bo'lgan konsepsiyalar qatoriga kirib ulgurdi. Adabiyot, ijtimoiy-maishiy, rasmiy, ilmiy sohalarda uchraydigan diskursning qo'llanish doirasi, lingvopragmatik, kognitiv xususiyatlari tahlil qilish yangi yondashuvlarni va qoidalarini talab qiladi.

Til ijtimoiy hodisa, insoniyat taraqqiyotining muhim omillaridan biri bo'lib, u nafaqat aloqa vositasi, balki jamiyatning siyosiy, ilmiy, ijtimoiy, madaniy holatini aks ettiruvchi muhim ko'zgu hisoblanadi. Har bir jamiyat o'zining ilmiy salohiyati va iqtisodiy-siyosiy sharoitiga qarab tilning rivojlanish yo'nalishini belgilaydi hamda undagi diskurslarni shakllantiradi. Xususan ilmiy diskursning shakllanishi va rivojlanishi ilm-fan, texnika va innovatsiyalar bilan bog'liq tarzda bo'lib, ilmiy anjumanlar, konferensiya va seminarlarda qilingan ilmiy ma'ruzalarda o'z aksini topadi. Ayni shu ma'noda diskurs muhim masala sifatida yashovchan va e'tibortalab tadqiqot obyekti sanaladi. Shu bois ilmiy diskurs masalasiga alohida e'tibor qaratish va uning mohiyatini anglab yetish muhim ahamiyatga ega bo'lmoqda. Jumladan, o'zbek ilmiy diskursining lingvistik xususiyatlari, janrlar masalasi, ilmiy diskursda matn muammosi masalalarini tadqiq qilish, diskurs tushunchasining intertekstuallik hodisasi bilan bog'liqligi, intertekstuallikni ifodalashda ilmiy diskurs strukturasi va lisoniy tizimini o'rganishga ehtiyoj tug'ilmoqda. Ilmiy diskurs muloqotning alohida turiga mansub bo'lib, u tinglovchiga kuchli ta'sir ko'rsatish xususiyatiga ega. Shu bois, ilmiy dalillash mexanizmlarini aniqlash g'oyat muhim ahamiyat kasb etadi. Shu nuqtayi nazardan, ilmiy diskursni o'rganishning ahamiyati ilmiy ma'lumotlarni asoslash va tekshirishning eng

maqbul yo‘llarini izlash zarurati bilan belgilanadi. Bu ma’lumotlar dunyoning ilmiy manzarasiga uyg‘un tarzda kiritilishi lozim.

Muloqot jarayonida lisoniy birliklar bilan bir qatorda nolisoniy (pragmatik, psixologik, ijtimoiy-madaniy) omillar ham mavjud bo‘lib, ularni o‘rganishga bo‘lgan qiziqish tilshunoslikda “diskurs” atamasini kengroq o‘rganilishga olib keldi. Diskurs va uning tahlili bo‘yicha ko‘pgina olimlar tomonidan ilmiy izlanishlar qilingan. Jahon tilshunosligida diskurs nazariyasi va uni tahlil qilish amaliyotining vujudga kelishi hamda rivojlanishi XX asrning 60-70-yillaridan boshlangan. Jumladan, N.D. Arutyunova, T.A. van Deyk, O.V. Aleksandrova, Ye.S. Kubryakova va A.Yu. Popovalarning diskursiv tilshunoslik sohasidagi tadqiqotlari; P.Ya. Galperin, T.A. Karasik, L.A. Axtayeva va M.P. Kotyurovalarning ilmiy diskurs xususiyatlari bo‘yicha izlanishlari; M.M. Baxtin, L.S. Vigotskiy, I.V. Arnold, Yu. Kristeva, J. Jennett, I.P. Smirnova, Ye.A. Bajenova, N.A. Kuzmina, F.G. Plett, N. Pyege-Gro va R. Bartlarning intertekstuallik sohasidagi ishlari; V.E. Chernyavskaya va Ye.V. Mixaylovalarning ilmiy diskurs xususiyatlari va intertekstuallik hodisasini o‘rganishga bag‘ishlangan ishlarida diskurs va intertekstuallik hodisalari haqida qimmatli ma’lumotlarni uchratamiz.

Diskurs atamasi lotin tilidan kirib kelgan bo‘lib, discursus- “harakat“, “aylanish“, “suhbat“, “muloqot“ degan ma’nolarni anglatadi. Tadqiqotlar shuni ko‘rsatadiki, diskurs lingvistik termin sifatida ishlatilgunga qadar asosan mantiqiy va epistemologik (falsafaning “bilim“ haqidagi sohasi) jihatdan ta’riflangan.

F. Sossyurning izdoshi fransuz lingvisti E. Benvenist fransuz lingvistik an’anasida umuman nutq, matn va terminologik ma’noni anglatuvchi “diskurs” so‘ziga birinchilardan bo‘lib “so‘zlovchi tomonidan o‘zlashtirilgan nutq” degan yangi ta’rifni berdi. U F. Sossyur tomonidan kiritilgan til sistema (langue) sifatida va til nutq sifatida (parole) oppozitsiyalarini ma’qullamadi va nutq (parole) so‘zining o‘rniga diskurs atamasini kiritdi. Olim til va diskurs birliklarini farqlagan ilk tilshunoslardan biri hisoblanadi. Uning fikricha “diskurs” gapning yuqori turadigan sathi bo‘lib, har doim voqea-hodisa sifatida namoyon bo‘ladigan so‘zlovchi nutqi orqali faollashadigan diskret (alohida) harakatdir [Godrich P., 1992: 135]. Diskurs nazariyasining uzoq ildizlarini M.M. Baxtin ishlarida ham ko‘rish mumkin.

Qator manbalarda “Diskurs” atamasi tilshunoslik sohasiga bir necha yo‘nalishlardan kirib kelganligi ta’kidlanadi: bir tomondan, uni 1952-yilda distributiv va transformatsion tahlil asoschisi Z. Xerris joriy etgan deb hisoblashadi. U “Diskurs tahlili” nomli maqolasida diskurs tahlili tushunchasini nutqni tahlil qilish usuli sifatida ta’kidlaydi va matnda mavjud bo‘lgan lingvistik elementlarni, matnning nutqiy vaziyatga munosabatini ko‘rib chiqadi [Harris Z., 1952: 30]. Olim diskurs tushunchasini nutqning ma’lum sohasi hamda maxsus leksikaga ega bo‘lgan nutqiy faoliyat bilan chambarchas bog‘liq hodisa sifatida tariflaydi. Boshqa tomondan esa, “diskurs” atamasi tilshunoslikka “kommunikativ harakat nazariyasi” muallifi Yurgen Xabermasning ishlari orqali “lingvistik burilish” deb atalgan jarayon davomida falsafa va sotsiologiyadan kirib kelganligi haqida qarashlar mavjud [Воркачев С, Воркачевaa E., 2019: 14].

O‘tgan asrning 60-yillarida til ijtimoiy voqelik sifatida qaralib, muloqotning shakllanish jarayonini o‘rganish boshlandi. Nutq nazariyasi pragmatik yo‘nalishni tanladi va matndan anglashilgan ma’no tahlil qilina boshladi. Diskurs tahlili va matn lingvistikasi tilshunoslikning bir-biriga yaqin, ba’zan esa tenglashtiriladigan sohalarini tashkil etgan. Biroq 70-yillarning oxiri va 80-yillarning boshlarida “matn” va “diskurs” tushunchalarining asta-sekin farqlanishidan kelib chiqadigan ularni chegaralash tendensiyasi paydo bo‘ldi. Matn deganda, asosan, mavhum, rasmiy tuzilma, diskurs deganda esa uning mental jarayonlar nuqtayi nazaridan va ekstralingvistik omillar bilan bog‘liq holda ko‘rib chiqiladigan turli xil aktuallashtirish turlari tushunildi va diskurs tahlili asosan tavsifiy va eksperimental usullarda amalga oshirildi.

Diskurs tahlili britaniyalik tilshunos M. Xollideyning tilga nisbatan funksional yondashuvlariga katta ta’sir ko‘rsatdi. M. Xollideyning ilmiy ishlari tilning ijtimoiy vazifalari, mazmuni, rasmiy yozma va og‘zaki nutq tahlilini o‘z ichiga oladi. Britaniya an’analari asosan badiiy adabiyot diskursiv tahlilga tortilgan. Britaniya ilmiy nazariyalari struktur tilshunoslik

tamoyillari, ya'ni til birliklarni ajratish, nutqning mukammal shakllangan birliklarini tavsiflovchi qoidalar majmui asosida qurilgan. Amerikalik tilshunoslarning diskursiv tahlili etno-metodik an'analarga, odamlarning tabiiy muloqotini kuzatish jarayonida ilmiy izlanishlarga asoslanadi. Praga tilshunoslik maktabi ham diskursiv tahlilning rivojlanishiga katta hissa qo'shdi, xususan, bu maktab grammatika va nutq o'rtasidagi o'zaro bog'liqlikni tahlil qilganligi alohida e'tirofga loyiqdir. Fransuz maktabi vakillari (A. Greymas, J. Derrida, Yu. Kristeva, M. Fuko) fikriga ko'ra, "diskursni -eng avvalo, muayyan ijtimoiy-siyosiy guruh yoki davrga xos bo'lgan ma'lum bir nutq turi" sifatida tahlil qildi [Greimas A. J., 1966: 121-123].

O'zbek tilshunosligida ham diskurs tushunchasi keng qamrovda o'rganilayotgan o'ziga xos fenomen hisoblanadi. Xususan, oxirgi 20 yillik davomida "diskurs" atamasining lingvistika bilan birgalikda boshqa ilmiy yo'nalishlarda ham aktualligi oshib borayotganligini kuzatishimiz mumkin. Zamonaviy tilshunoslikda til tizimini antroposentrik nuqtayi nazardan o'rganish, asosan, kognitiv tilshunoslik, lingvistik semantika, sotsiolingvistika, psixolingvistika, pragmalingvistika hamda lingvokulturologiyaga oid tadqiqotlarda amalga oshiriladi. Olimlar diskurs atamasini ushbu sohalar bilan bog'liq tarzda ko'plab dissertatsiya, monografiya, darslik, qo'llanma hamda maqolalarda izohlab o'tganlar. XXI asrning dastlabki yillaridanoq o'zbek tilshunosligida antroposentrik paradigma asosida qilingan tadqiqotlar monografik planda namoyon bo'la boshladi. Tilshunos S.M.Mo'minovning "O'zbek muloqot xulqining ijtimoiy-lisoniy xususiyatlari" mavzusidagi izlanishlarida o'zbek xalqining o'ziga xos muloqot xulqini ijtimoiy-lisoniy xususiyatlarini tahlilga tortilgan. N.Mahmudovning tilshunoslikda antroposentrik paradigmaning shakllanishi haqidagi izlanishlari, M. Xolmurodovanning sotiopragmatikaga oid tadqiqotlari, D.Xudoyberganovanning badiiy matnning antroposentrik, xususan, kognitiv-semantik, psixolingvistik va lingvokulturologik tadqiqiga bag'ishlangan izlanishlari, Sh.Bobojonovanning ta'limiy diskur bo'yicha izlanishlari, Sh.Safarov, M.Hakimovlarning pragmalingvistika bo'yicha qilgan tadqiqotlari, Sh. Iskandarova, G. Odilova, A. Pardayev, L. Raupova, N.Xusanov, D. Shonazarova va boshqalarning ishlarida tadqiqot mavzusi bilan bog'liq qimmatli ma'lumotlar uchraydi

Diskurs tushunchasi abstrakt va polesemantik tushuncha bo'lib, asosan nutqiy jarayonning aniq bir vaqt va makonda sodir bo'lishini talab qiladi va har qanday og'zaki yoki yozma nutq diskurs bo'lishi uchun nutqiy harakat amalga oshirilishi kerak. Diskurs tipologik jihatdan xilma-xil bo'lib, turli nuqtayi nazarlardan o'rganiladi va bir qator turdosh fanlarning tadqiqot obyekti sanaladi. Masalan, matn tilshunosligida diskurs til, nutq, matn, dialog, uslub kabi an'anaviy tushunchalar bilan bog'lanadi; lingvokulturologiyada ma'lum bir etnosga xos diskursning o'ziga xos xususiyatlari aniqlanadi; pragmatikada diskurs muloqot ishtirokchilarining nutqiy-fikrlash faoliyati (og'zaki va og'zaki bo'lmagan ifoda) sifatida ko'rib chiqiladi; kognitologiyada diskurs konseptlar, freymlar, ssenariylar va boshqalar ko'rinishida o'rganiladi; sotsiolingvistikada kommunikantlar - jamiyatning ma'lum bir ijtimoiy guruh vakillari o'rtasidagi turli xil muloqot turlari o'rganiladi, muayyan ijtimoiy guruh vakillari (texnik, iqtisodiy, tibbiy va shunga o'xshash diskurslar) yoki keng ijtimoiy doiradagi (siyosiy, diniy va boshqa diskurslar) kommunikantlar o'rtasidagi turli xil muloqot turlari tadqiq qilinadi. Shu munosabat bilan zamonaviy tilshunoslikda diskursni o'rganishning kommunikativ, pragmatik, struktur-sintaktik, sotsiolingvistik kabi asosiy yondashuvlari ajratib ko'rsatiladi

Ilmiy nutq yaqqol ifodalangan matnlararo bog'liqlik va diskursivlik bilan ajralib turadi. Shu sababli, ilmiy fikrlarni to'g'ri tushunish va ularni yaratish uchun ko'pincha umumiy qoidalarga, oldingi matnlarga va obro'li shaxslar, olimlar nomlariga va ularning nazariyalariga tayanish xosdir. Zamonaviy tilshunoslikda intertekstuallik hodisasi muhim va dolzarb lingvistik muammolardan biri sanaladi. Chunki u matnlar - keng ma'noda diskurslar va madaniyatlar o'rtasidagi o'zaro ta'sir bilan bog'liq. Shu jihatdan ham diskurs tushunchasi intertekstuallik hodisasi bilan bog'liq hisoblanadi. Har bir tadqiqot asosida muayyan muammoga turli xil yondashuv va muallifning nuqtayi nazari yotadi. Mavjud yondashuvlarning xilma-xilligiga qaramay (funksional-stilistik, kommunikativ-pragmatik, sotsiolingvistik, lingvomadaniy,

kognitiv-semantik, pragmalingvistik, tarixiy, lingvokognitiv), ilmiy diskursning barcha tadqiqotlari zamonaviy akademik muhitda ilmiy muloqot va ilmiy bilimni uzatish diskurs orqali amalga oshiriladi.

Foydalanilgan adabiyotlar:

1. Goodrich, P. (1992). Legal discourse: studies in linguistics, rhetoric and legal analysis Macmillan. (p. 135).
2. Harris Z.S Discourse analysis //Language, 1952. –30
3. Ismailov, A., et al. “A comparative study of stemming algorithms for use with the Uzbek language” 2016 3rd International conference on computer and information sciences (ICCOINS). IEEE, 2016.
4. Khusainov, A., Suleymanov, D., Gilmullin, R., Minsafina, A., Kubedinova, L., & Abdurakhmonova, N. (2020, November). First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems. In CMCL (pp. 90-101).
5. Isroilov, J., & Abdurakhmonova, N. (2018). Personal names spell-checking—a study related to Uzbek. Journal of Social Sciences and Humanities Research, 6(02), 1-6.
6. Воркачев С.Г., Воркачева Е.А. Дискурс и его типология в российской лингвистике // Актуальные проблемы филологии и педагогической лингвистики. 2019; 3: 14-21 <https://philjournal.ru/upload/2019-3/14-21.pdf>
7. A. J. Greimas, *Sémantique structurale. Recherche de méthode*. 1966: 121-123
8. IBODULLAYEVA, K. (2025). INTERTEXTUALITY IN SCIENTIFIC DISCOURSE (EXAMPLE OF PHILOLOGICAL ARTICLES IN UZBEK) .News of the NUUz, 1(1.12.1),185-187. <https://doi.org/10.69617/nuuz.v1i1.12.1.5744>
9. <http://rus-yaz.niv.ru/doc/dictionary/linguistic-encyclopedic/articles/405/diskurs.htm>
10. <https://interscience.uz/index.php/home/article/download/3578/3206/8512>
11. https://dictionary.cambridge.org/dictionary/english/discourse#google_vignette
12. <https://www.collinsdictionary.com/sentences/english/discourse>

MASHINASOZLIK TERMINOLOGIYASINING SHAKLLANISH BOSQICHLARI

Maxsudova Xolisxon Ummatovna

Andijon davlat texnika instituti

Tillar va gumanitar fanlar kafedrası dotsenti

E-mail: xolisxon0408@gmail.com

Annotatsiya: Ushbu maqolada mashinasozlik terminologiyasining tarixiy taraqqiyot bosqichlari, uning lingvistik shakllanishi, milliy texnik tafakkur bilan bogʻliqligi va zamonaviy terminlar tizimining shakllanishiga taʼsir etuvchi omillar tahlil qilinadi. Soha terminologiyasi rivojida ilm-fan, sanoat va tarjimashunoslikning oʻzaro aloqasi alohida oʻrganiladi.

Kalit soʻzlar: mashinasozlik, terminologiya, texnik til, tarixiy taraqqiyot, soha leksikasi, tarjima

Abstract: This article analyzes the stages of historical development of mechanical engineering terminology, its linguistic formation, its connection with national technical thinking, and the factors influencing the formation of a modern terminological system. The relationship between science, industry, and translation studies in the development of industry terminology is studied separately.

Keywords: mechanical engineering, terminology, technical language, historical development, industry lexicon, translation

Аннотация: В статье анализируются этапы исторического развития терминологии машиностроения, ее языковое становление, ее связь с национальной технической мыслью, а также факторы, повлиявшие на формирование современной терминологической системы. При разработке отраслевой терминологии отдельно изучается взаимосвязь науки, промышленности и переводоведения.

Ключевые слова: машиностроение, терминология, технический язык, историческое развитие, отраслевая лексика, перевод

Texnik sohalar, jumladan, mashinasozlik tili inson tafakkurining yuqori darajadagi ixtisoslashgan shaklidir. Ushbu tilning asosiy komponenti boʻlgan **terminologiya** texnik tafakkur va amaliy faoliyatda hal qiluvchi oʻrin tutadi. Har bir texnik soha kabi, mashinasozlik tili ham oʻziga xos shakllanish yoʻlini bosib oʻtgan. Quyida oʻzbek tilida mashinasozlik terminologiyasining shakllanish bosqichlari va ularning xususiyatlari tahlil qilinadi

1. Ilk bosqich – terminlarning xalq ogʻzaki nutqdan kirib kelishi (XIX asr oxiri – XX asr boshi)

Bu davrda **sanoatlashtirish jarayonining boshlanishi** bilan birga xorijiy (asosan rus va fransuz) tillardan oʻzbek tiliga texnik atamalar oʻzlashgan. Terminlarning koʻpchiligi fonetik moslashuv yoʻli bilan kirib kelgan:

- **Stanok** → stanok
- **Motor** → motor
- **Mashina** → mashina

Bu davrda terminlar adabiy meʼyorga solinmagan, leksik jihatdan chalkash boʻlgan va koʻpincha norasmiy ishlatilgan.

2. Sovet davri – terminlarni normativlash va tizimlashtirish bosqichi (XX asr 30–80-yillari)

Sovet Ittifoqi davrida texnik bilimlar jadallik bilan rivojlandi, yangi mashinalar, mexanizmlar, stanoklar va texnologiyalar paydo boʻldi. Bularning nomlarini ifodalovchi **terminlar markazlashgan holda rus tilidan tarjima qilinib**, koʻplab ixtisoslashgan lugʻatlar tuzildi.

- O'zlashtirilgan terminlar: reduktor, frezer, tokarka
- So'z yasash orqali: kesuvchi asbob, aylanish momenti

Bu davrda **terminologik komissiyalar** tashkil etilgan, lekin o'zbek tilidagi tarjimalar ko'pincha rus tilidan to'g'ridan-to'g'ri tarjima qilinardi.

3. Mustaqillikdan keyingi bosqich – milliylashuv va modernizatsiya (1991 yildan hozirgacha)

O'zbekiston mustaqillikka erishgach, texnik terminlar tizimini **milliylashtirish** tendensiyasi kuchaydi. Terminlarni rus tilidan tarjima qilish o'rniga, o'zbek tilidagi muqobillarni yaratish harakatlari boshlandi:

- Stanok → arga mashinasi
- Shpindel → markaz aylanish mili
- Privod → harakatlantiruvchi uzatma

Shuningdek, zamonaviy texnologiyalar (kompyuterlashtirish, avtomatlashtirish) bilan bog'liq **inglizcha terminlar** to'g'ridan-to'g'ri kirib kelmoqda:

- CNC (Computer Numerical Control) – kompyuter boshqaruvli dastgoh
- Mechatronics – mexatronika

4. Hozirgi davr – raqamli transformatsiya va ko'p tillilik davri

Bugungi kunda texnik terminlar:

- Ingliz tilidan kirib kelmoqda, lekin rus tilidagi ekvivalentlar ham saqlanib qolmoqda.
- Sun'iy intellekt va avtomatik tarjima tizimlari (Google Translate, ChatGPT, DeepL va boshqalar) orqali tarqalmoqda.
- Texnik universitetlar, korxonalar va sanoat vakillari tomonidan shakllantirilmoqda.

Terminlarning standartlashtirilgan, ko'p tilli korpuslariga ehtiyoj ortmoqda. Bu esa tilshunoslar, tarjimonlar, IT mutaxassislar hamkorligini talab qiladi.

Mashinasozlik terminologiyasining shakllanishi uzluksiz tarixiy jarayondir. U xalq tilidan texnik standartlarga o'zgarib bordi. Hozirgi bosqichda bu terminologiya milliy til taraqqiyoti, xalqaro integratsiya va raqamli texnologiyalar bilan chambarchas bog'liq holda rivojlanmoqda. Uning tizimli tadqiqi va zamonaviy usullar bilan kodifikatsiyasi kelajakda texnik bilimlar uzatishda muhim omil bo'lib qoladi.

Foydalanilgan adabiyotlar ro'yxati:

1. Axmedov A. "O'zbek tilida texnik terminlarning shakllanishi va ishlatilishi". – Toshkent: Fan, 2005.
2. Xasanov N. "Texnik terminlar tarjimasi va ularning o'zbek tilida ifodalanishi". – Toshkent: O'zMU, 2012.
3. Zohidov R. "Terminologiyaning nazariy asoslari". – Toshkent: O'zR FA nashriyoti, 2008.
4. Jumanioyov Sh. "Soha terminologiyasi: tuzilish, tarjima va zamonaviy yondashuvlar". – Termiz: Surxon, 2020
5. Mirtojiyev M., Salomov G. "O'zbek tilining leksikologiyasi". – Toshkent: O'zMU, 2015.
6. Xolmatov B. "Mashinasozlik va metallga ishlov berish atamaları lug'ati". – Toshkent: Texnika, 1991
7. Safarov I. "Texnik matnlar tarjimasi: nazariya va amaliyot". – Toshkent: Fan, 2018.
8. Qodirov A. "Texnik terminlarning korpus lingvistikasi asosida tahlili". – O'zMU ilmiy maqolalar to'plami, 2022
9. Крысин Л.П. "Современные проблемы русской терминологии". – Москва: Наука, 2004.
10. Кабакчи В.В. "Теория и практика перевода научно-технической литературы". – СПб: Издательство СПбГУ, 2010.
11. Лотте Д.С. "Основы построения научно-технической терминологии". – Москва: Наука, 1961.

12. Гарбовский Н.К. "Теория перевода". – Москва: Издательство Московского университета, 2007.
13. ISO 704:2009 – "Terminology work — Principles and methods"
14. ISO 1087-1:2000 – "Terminology work — Vocabulary — Part 1: Theory and application"
15. Pearson J. "Terminology in Translation: A Model for Training and Practice". – Amsterdam: John Benjamins, 1998.
16. Cabré M.T. "Terminology: Theory, Methods and Applications". – Amsterdam: John Benjamins, 1999.
17. Bowker L. "Computer Applications in Terminology: Tools for Terminographers and Translators". – Amsterdam: Benjamins, 2003.
18. Drouin P. "Term extraction using non-contiguous word sequences". In: Terminology, Vol. 9(2), 2003.
19. Temurshoev A. "Mashinasozlik lugʻatlari va ularning oʻzbek tilidagi variantlarini standartlashtirish". – Texnik tillar konferensiyasi materiallari, 2021.
20. Oʻzbek tilining izohli lugʻati. 5 jildlik. – Toshkent: Oʻzbekiston Milliy ensiklopediyasi, 2006–2008.

O'ZBEK TILINING MORFOLOGIK ANALIZATORINI TADQIQ ETISH VA YARATISH

Shaydullayeva Muxlisa
Fazliddinova Mavluda
Ismanqulova Xumora
Mamarayimova Aziza

Toshkent davlat sharqshunoslik universiteti
Sharq xalqlari tillari va adabiyoti instituti
tarjimashunoslik va jurnalistika oliy maktabi filologiya fakulteti
kompyuter lingvistikasi guruhi talabalari

Annotatsiya: Ushbu ishda o'zbek tili uchun morfologik tahlil usullari o'rganiladi va tegishli analizator ishlab chiqiladi. O'zbek tili agglutinativ turkiy til sifatida murakkab morfologik tuzilishga ega bo'lib, bu avtomatik qayta ishlash uchun maxsus yondashuvlarni talab qiladi. Tadqiqotning dolzarbligi o'zbek tili uchun NLP vositalarining yetarli darajada ishlab chiqilmaganligi bilan belgilanadi. Ishning maqsadi mavjud usullarni tahlil qilish, morfologik analizatorni yaratish va uning samaradorligini baholashdan iborat. Natijalar mashina tarjimasi, qidiruv tizimlari va boshqa NLP masalalarida qo'llanilishi mumkin.

Kalit so'zlar: Natural Language Processing, mashina tarjimasi, morfologik analiz.

Kirish

Zamonaviy davrda matnli ma'lumotlar hajmining keskin oshib borishi ularni avtomatik tarzda qayta ishlash ehtiyojini yuzaga keltirmoqda. Mazkur ehtiyoj matnni tahlil qilishning avtomatlashtirilgan vositalarini ishlab chiqishga sabab bo'ldi. Bunday vositalar inson faoliyatining turli sohalarida keng qo'llanila boshlagan bo'lib, ularning ishlash tezligi va aniqligi muhim funksional xususiyatlar sifatida ajralib turadi.

Tabiiy tilning murakkab tuzilishi uni tahlil qilish jarayonini bir nechta bosqichlarga ajratishni talab etadi. Ushbu bosqichlardan biri — morfologik tahlil bosqichidir. Morfologik tahlil bosqichi matndagi so'z birliklarini ajratish, ularning nutqdagi grammatik rollarini aniqlash hamda morfologik xususiyatlarini belgilash uchun xizmat qiladi. Ushbu bosqichda olingan ma'lumotlar matnni chuqurroq tahlil qilish uchun asos bo'lib, ayniqsa, keyingi sintaktik tahlil bosqichida matnning sintaktik tuzilmasini aniqlashda muhim rol o'ynaydi [1].

Zamonaviy tabiiy tilni qayta ishlash texnologiyalari (Natural Language Processing, NLP) axborot tizimlari, mashina tarjimasi, qidiruv algoritmlari va sun'iy intellektning boshqa sohalarini rivojlantirishda muhim rol o'ynaydi. Lingvistik tahlilning asosiy bosqichlaridan biri morfologik tahlil bo'lib, u so'zning grammatik xususiyatlarini (so'z turkumi, kelishik, son, zamon va boshqalar) aniqlashga imkon beradi. Ingliz yoki rus tili kabi keng tarqalgan tillar uchun yaxshi ishlab chiqilgan morfologik analizatorlar mavjud, ammo o'zbek tili uchun bunday vositalar yetarlicha o'rganilmagan va qo'shimcha tadqiqotlarni talab qiladi.

Turkiy tillar oilasiga mansub o'zbek tili boy morfologiya, agglutinativ tuzilish va so'z yasash xususiyatlariga ega bo'lib, avtomatik qayta ishlashda ma'lum qiyinchiliklarni keltirib chiqaradi. Turkiy tillar uchun NLPga qiziqish ortib borayotganiga qaramay, o'zbek tili uchun samarali morfologik analizatorni ishlab chiqish dolzarb vazifa bo'lib qolmoqda.

O'zbek tili, flektiv va derivatsion qo'shimchalarga ega bo'lgan, agglutinativ morfologik tizimga asoslangan til hisoblanadi. Bu tilning asosiy xususiyatlaridan biri — so'zlarning morfologik shakllarining ko'pligi va ularning ketma-ket qo'shilishi orqali yangi ma'nolarni ifodalash imkoniyatidir. O'zbek tilining morfologik strukturasi juda boy va keng qamrovli bo'lib, so'zlar bir necha ma'lumotni o'z ichiga olishi mumkin. Bu ma'lumotlar quyidagilarni o'z ichiga oladi:

Egalik: Soʻzga qoʻshilgan qoʻshimchalar yordamida egalik (kimga tegishliligi) maʼlumoti beriladi.

Koʻplik/Birlik: Soʻzlarning koʻplik va birlik shakllari orqali son maʼlumoti ifodalanadi.

Kelishik: Oʻzbek tilida kelishik oʻziga xos oʻzgarishlarni koʻrsatadi va bu tilda keng tarqalgan til hodisasi hisoblanadi.

Ushbu morfologik elementlar, koʻplab flektiv shakllarning mavjud emasligi va maʼlum bir grammatik maʼlumotlarning oʻzgarishi tufayli, mashina tarjimasida murakkabliklar yuzaga kelishi mumkin. Oʻzbek tilidagi agglutinativ morfologiyaning oʻziga xos xususiyatlari, masalan, bir soʻzning bir nechta qoʻshimchalarni qabul qilish imkoniyati, morfologik analizator yaratish jarayonida alohida eʼtiborga olinishi kerak [2].

Morfologik analizatorni ishlab chiqishda oʻzbek tilining flektiv va derivatsion qoʻshimchalariga alohida urgʻu berish, ularning toʻgʻri ishlatilishi va ketma-ket qoʻshilishi tizimini tushunish zarur. Bu, oʻz navbatida, mashina tarjimasida tizimining yuqori sifatini taʼminlashga yordam beradi.

Morfologik analizator muayyan tokenning morfologik tarkibini tahlil qilish uchun masʼul boʻlgan dastur sifatida izohlanadi. Morfologik analizator berilgan tokenni tahlil qiladi va turkum, turli grammatik maʼnolar kabi maʼlumotlarni shakllantiradi.

Morfologik analizator yaratishda quyidagilar tushunchalarni oʻrganish maqsadga muvofiq:

- 1)Morfologik birliklar;
- 2)Stemmer;
- 3)Lemmatizatsiya;
- 4)Oʻzbek tilida stemmerni ishlab chiqish;
- 5)Morfologik analizator;
- 6)Morfologik generator;
- 7)Qidiruv tizimini ishlab chiqish [3].

Matnlarni tahlil qilishda eng yuqori aniqlik va keng imkoniyatlarga erishish toʻliq (kompleks) tahlilni amalga oshirish orqali mumkin boʻladi. Biroq, bunday tahlilni yaratish jarayonida yuzaga keladigan murakkabliklar shunchalik jiddiyki, amaliyotda hozirgi vaqtgacha ishlab chiqilgan barcha nazariy tamoyillar toʻliq amalga oshirilmagan. Ushbu sohadagi asosiy muammolardan biri — affikslar (qoʻshimchalar)ning nihoyatda koʻp kombinatsiyalari bilan ishlashdagi murakkablikdir [4].

#oddiy morfologik analiz

import re

def simple_morph_analysis(word):

Oddiy qoʻshimchalar roʻyxati

suffixes = ["lar", "ni", "ga", "dan", "lik", "chi", "larimiz", "ingiz", "ning"]

for suffix in suffixes:

if word.endswith(suffix):

root = word[:-len(suffix)]

return root, suffix

return word, None

Sinov uchun soʻzlar

words = ["kitoblar", "oʻquvchi", "maktabimizdan", "doʻstingizga"]

for word in words:

root, suffix = simple_morph_analysis(word)

print(f'Soʻz: {word} | Ildiz: {root} | Qoʻshimcha: {suffix}')

Kiritilgan soʻzlar va natijalar:		
"kitoblar"	Ildiz	kitob
	Qoʻshimcha	"lar" (koʻplik qoʻshimchasi)

"o'quvchi"	Ildiz	o'quv
	Qo'shimcha	"chi" (sifat yasovchi qo'shimcha)
"maktabimizdan"	Ildiz	maktabimiz
	Qo'shimcha	"dan" (kelishik qo'shimchasi)
"do'stingizga"	Ildiz	"do'stingiz"
	Qo'shimcha	"ga" (yo'nalish kelishigi qo'shimchasi)

O'zbek tili uchun morfologik analizatorni ishlab chiqish NLP sohasidagi dolzarb muammolardan biridir. Ushbu tadqiqot natijalari nafaqat lingvistik jihatdan, balki amaliy ahamiyatga ham ega bo'lib, mashina tarjimasi, matn tahlili va sun'iy intellekt sohalarida qo'llanilishi mumkin. Kelajakda ushbu yo'nalishda chuqurroq tadqiqotlar olib borish, shuningdek, morfologik analizatorni yanada takomillashtirish maqsadga muvofiq bo'ladi.

Adabiyotlar ro'yxati:

1. Mohbey K. K., Tiwari S. Preprocessing and morphological analysis in text mining // International Journal of Electronics Communication and Computer Engineering. - 2011. - Vol. 2. - № 2. - P. 1-7.
2. Бобур Аллабердиев, Шарипбай Алтынбек, Зафар Рахмонов "Моделирование аналитических форм глагола в узбекском языке как этап морфологического анализа в машинном переводе" Национальный университет Узбекистана.
3. Botir Elov, Shahlo Hamroyeva, Dirlabo Elova "Morfologik analizatorni yaratish usullari" Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti.
4. Абжалова М. Принципы создания морфологического модуля для программы автоматического анализа узбекского текста

MASHINA TARJIMASI UCHUN TARJIMA XOTIRASINI YARATISH TAMOYILLARI

Shamsiyeva Gulshoda Asliddin qizi,

Mirzo Ulugʻbek nomidagi

Oʻzbekiston Milliy universiteti tadqiqotchisi

E-mail: shamsiyeva_g@nuu.uz

Annotatsiya: Ushbu maqolada mashina tarjimasi uchun muhim jarayonlardan biri boʻlgan tarjima xotirasini yaratish, uning nazariy asoslari haqida soʻz boradi. Mashina tarjimasida terminlar va iboralar tarjima xotirasidan foydalanish orqali tarjimaning butun matni boʻylab terminologiyaning izchilligi taʼminlanishiga ham toʻxtalangan. Shuningdek, tarjima xotirasi dasturlarini guruhlariga ajratish mezonlari va tarjima xotirasi orqali tarjima qilish bosqichlari bilan birga TM ning asosiy xususiyatlari, afzalliklari va cheklovlari ilmiy manbalarga asoslangan holda tahlil qilinadi.

Kalit soʻzlar: tarjima xotirasi, tarjima xotirasi maʼlumotlar bazasi, online TM dasturlari, mahalliy TM dasturlari, matn segmenti, parallel korpus.

Annotation: This article discusses the creation of translation memory, one of the important processes for machine translation, and its theoretical foundations. In machine translation, the consistency of terminology throughout the entire text of the translation is ensured through the use of translation memory of terms and phrases. Also, along with the criteria for dividing translation memory programs into groups and the stages of translation through translation memory, the main features, advantages, and limitations of TM are analyzed based on scientific sources.

Keywords: translation memory, translation memory tools, online TM programs, local TM programs, text segment, parallel corpus.

Аннотация: В данной статье рассматривается создание памяти перевода, которая является одним из важных процессов для машинного перевода, и ее теоретические основы. Также было отмечено, что в машинном переводе термины и фразы обеспечивают последовательность терминологии на протяжении всего текста перевода за счет использования памяти перевода. Также на основе научных источников проанализированы основные характеристики, преимущества и ограничения ТМ, наряду с критериями группировки программ памяти перевода и этапами перевода через память перевода.

Ключевые слова: память перевода, база данных памяти перевода, онлайн-программы ТМ, локальные программы ТМ, сегмент текста, параллельный корпус.

Tarjima xotirasi (Translation Memory, TM) – bu kompyuter yordamida tarjima (Computer-Aided Translation, CAT) jarayonida tarjimonlarga yordam berish uchun ishlab chiqilgan texnologiya boʻlib, avval tarjima qilingan segmentlarni (soʻzlar, iboralar, jumlar yoki paragraflar) saqlaydi va takror ishlatishga imkon beradi. TM tizimlari tarjima jarayonining samaradorligini oshirish, izchillikni taʼminlash va vaqtni tejash uchun qoʻllaniladi. Lynne Bowker tomonidan yozilgan “Computer-Aided Translation Technology: A Practical Introduction” [Lynne Bowker., 2002] asarida TM tizimlarining ishlash tamoyillari, afzalliklari va cheklovlari haqida batafsil maʼlumot berilgan. Tarjima xotirasi tizimlari eng soʻnggi rivojlanayotgan texnologiya ekanligi va ularning tarjimonlar orasida mashhurligi tobora ortib borayotgani aniq koʻrsatilgan. Ushbu tizimlar manba va maqsadli matnlarni bogʻlaydi va ushbu parallellarni maʼlumotlar bazasida saqlaydi. Bu tarjimonga avval tarjima qilingan parchalarni matnning boshqa oʻrinlarida yana ishlatish imkonini beradi. TM tizimi yangi boshlangʻich matnni avvalgi tarjimalar bilan taqqoslaydi. Ushbu parchalarning mosligi aniq moslik, toʻliq moslik, noaniq moslik, atamalar mosligi yoki kichik segmentlar mosligi boʻlishi mumkin. Ushbu texnologiyadan foydalanish

yangilanayotgan, qayta ko‘rib chiqilayotgan va takrorlanuvchi mazmunga ega bo‘lgan matnlar bilan cheklanadi. Shuningdek, bir mavzu doirasidagi qator matnlarga nisbatan ham qo‘llanilishi mumkin.

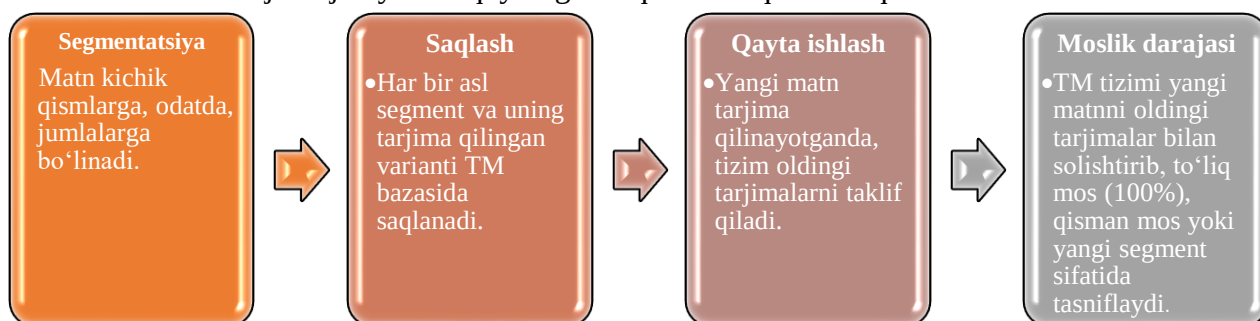
Tarjima jarayoni matn bilan ishlashda multifunksional faoliyat bo‘lib, matnga ijtimoiy va aniq fanlar nuqtai nazaridan yondashishni talab etadi. Tarjimada qo‘llanuvchi matnlar mavzusi ilmiy jihatdan keng bo‘lib, kognitiv fanlar, xususan, tarix, adabiyotshunoslik, o‘lkashunoslik kabi fanlar bilan bir qatorda tarix, fizika, qurilish, ekologiya, mexanika matematika kabi ilmiy doirasi texnika bilan bog‘liq fanlarga aloqador matnlar bilan ham ish ko‘radi.

Tarjima inson hayotining barcha sohalariga ta’sir etuvchi faoliyat bo‘lib, TM (Translation Memory) ni yo‘lga qo‘yish uchun kategorik-kontseptual apparat sifatida individual leksikani tayyorlash, soha uchun individual lug‘atni to‘plash va saqlash lozim bo‘ladi. Bunda TM uchun bazani shakllantirishda turli mavzulardagi turli muammolar haqidagi juda ko‘p matnlarni yig‘ish va ular bilan ishlashga ehtiyoj tug‘iladi.

Axborotni saqlash va eslab qolishda inson imkoniyatlari kompyuternig ma’lumotlar omboridan sust ekanligi ma’lum. Bundan tashqari, kerakli kontekstni, allaqachon tarjima qilingan matnning ma’lum bir qatorini, an’anaviy analog bosma versiyadagi lug‘atlarni qidirish bir xil ish sifatini saqlab, tarjimon ishining tezligiga ta’sir qiladi. Ushbu muammoni hal qilish va tarjima jarayonini optimallashtirish uchun 1985-yildan boshlab tarjima xotirasi yoki tarjima xotirasi ma’lumotlar bazalari (Translation Memory Tools – TMT, Translation Memory – TM) kabi maxsus axborot texnologiyalari paydo bo‘la boshladi. Bundan tashqari, SDL Trados, Déjavu kabi tarjima xotiralari berilgan tematik guruhlar uchun terminologiyani saqlashga hissa qo‘shadigan lug‘atlar va atamalarni saqlash imkonini beradi. Natijada, tarjima loyihasi doirasida ko‘p sonli matnlar bilan ishlashda o‘xshash jumalarni yoki bir xil atamalarni qayta tarjima qilishning hojati bo‘lmaydi, tarjimada allaqachon saqlangan tarjimalar ma’lumotlar bazasidan kerakli segmentni olishning o‘zi yetarli bo‘ladi, shu tariqa tizim ishni tezlashtiradi va optimallashtiradi. Terminlar bazasida lug‘atda mavjud iboralar, so‘zlar va atamalarning aniq matnlarini saqlash orqali tarjimaning butun matni bo‘ylab terminologiyaning izchilligi ta’minlanadi, bu ayniqsa tarjimonlar guruhi uchun tarjima loyihasi ustida ishlashda foydalidir. Bitta tarjima bazasi va atama bazasidan foydalanish turli tarjimonlarning tarjimalarida natijaning bir xilligini yo‘qotish ehtimolini sezilarli darajada kamaytiradi. Tarjima rivojlanishining hozirgi bosqichida CAT qurilmalaridan foydalanish masalasi tobora ommalashib borayotgan mavzuga aylanib bormoqda va uning dolzarbligi ortmoqda:

- Kompyuterga asoslangan tarjima (Computer Assisted Translation (CAT))
- Tarjima xotirasi (Translation Memory)
- Tarjima muharriri (Translation editors)
- Terminologik ma’lumotlar bazasi

TM tizimi tarjima jarayonini quyidagi bosqichlar orqali boshqaradi:



Bowker [Lynne Bowker., 2002] TM tizimlarining bu afzalliklarini ta’kidlab, ularning tarjimonlar uchun qulay vosita ekanligini qayd etadi.

TM tizimining asosiy afzalliklari quyidagilardan iborat:

- **Tezlik** – Takrorlanuvchi matnlar tezroq tarjima qilinadi.
- **Izchillik** – Terminologiya va iboralarining bir xil tarjima qilinishini ta’minlaydi.

- **Narxni kamaytirish** – Takroriy tarjimalar uchun vaqt va mablag‘ iqtisod qilinadi.
- **Tarjimonlar uchun qulaylik** – Avvalgi tarjimalarni ko‘rish va tahrirlash imkoniyati mavjud.

TM dasturlarini bir nechta mezonlarga ko‘ra ajratish mumkin:

1. Texnik amalga oshirishga ko‘ra: a) online TM dasturlari; b) mahalliy TM dasturlari. Online TM dasturlarida ish internet-brauzer oynasida amalga oshiriladi. Google Translator Toolkit va Wordfast Anywhere dasturlari shular jumlasidandir. Shu bilan birga, Google Translator Toolkit’dagi ish natijalari (tarjima xotirasi, lug‘atlar) boshqa foydalanuvchilarga ham taqdim etilishi mumkin. Mahalliy dasturlar mustaqil dasturlar yoki o‘rnatilgan ilovalar bo‘lishi mumkin. O‘rnatilgan ilovalar matn muharririda (masalan, Microsoft Word) o‘rnatiladi. Tarjima bevosita matn muharririda amalga oshiriladi. Ushbu yondashuvning afzalligi shundaki, foydalanuvchi tanish muharrir interfeysi bilan ishlaydi, unda matnni tahrirlashning variantlari ham mavjud. Bunday dasturlarning kamchiligi shundaki, unda fayl formatlarini qo‘llab-quvvatlash cheklangan bo‘ladi, odatda faqat matn muharriri formatlari mavjud bo‘ladi. Alohida ilovalar bilan ishlashda butun tarjima jarayoni ilova muhitida amalga oshiriladi.

2. Qo‘shimcha funksiyalarning mavjudligiga ko‘ra. Terminologiyaga oid dasturlar ko‘pincha TM dasturlariga o‘rnatiladi. Ular terminologiya ma‘lumotlar bazalarini almashish, terminologiyani saqlash va qayta ishlash imkonini beradi. Bu tarjimada terminologiyaning birligini ta‘minlash, muayyan korxona yoki subyekt doirasida til siyosatiga rioya etilishini ta‘minlash imkonini beradi. Tarjima xotirasini yaratish uchun ushbu juftlarni dastur ma‘lumotlar bazasiga keyingi kiritish uchun dastlabki matn va tarjimaning segmentlarini taqqoslash kerak bo‘ladi. Shuning uchun ko‘pincha TM dasturlarida matn segmentlarini (ko‘pincha jummalarni) birlashtirish funktsiyasi mavjud bo‘ladi.

Bundan tashqari, bevosita tarjima jarayonida leksik birliklarning noaniqligini bartaraf etish uchun ma‘lum bir so‘zning kontekstda qo‘llanilishiga qarash kerak bo‘ladi. TM dasturlari qoida tariqasida, so‘z moslashuvini yaratish qobiliyatini ta‘minlaydi - kontekstda so‘zning barcha qo‘llanishlari ro‘yxatini topish, shu bilan birga har bir holatda leksik birlikni tarjima qilish usullarini taqdim etadi. Aslida, bu funktsiya matnlarning ko‘p tilli korpusiga tegishli bo‘lib, leksik birliklarni qidirish matn korpusi bilan ishlashda ishlatiladigan interfeyslarga o‘xshash interfeyslar orqali amalga oshiriladi. O‘zaro muvofiq interfeysga ega til juftliklari uchun ba‘zi onlane parallel korpuslar mavjud. Masalan, TransSearch, WeBiText, TradooIt, SynchroTerm ni misol qilish mumkin. Axborot resursini qidirish tarjimon uchun lingvistik ma‘lumotlarning yetarlicha kuchli manbai bo‘lganligi sababli, qidiruv vositasiga ega parallel matn korpuslari ba‘zan tarjima xotirasining alohida turi sifatida ajralib turadi va tarjima xotirasi dasturlariga to‘g‘ri qarama-qarshi bo‘ladi, unda takrorlanuvchi segmentlarni qidirish ba‘zi matnlar uchun avtomatik ravishda amalga oshiriladi.

3. Dasturiy ta‘minotning umumiy mezonlariga ko‘ra: a) bepul; va b) umumiy dasturiy ta‘minot. Ular turli xil operatsion tizimlarni qo‘llab-quvvatlaydi. Masalan, STAR Transit - bu Shareware va faqat Windowsni qo‘llab-quvvatlaydi, OmegaT bepul va ochiq manba bo‘lib Windows, Linux, Mac OS ni qo‘llab-quvvatlaydi.

TM dasturlardan foydalanish tarjimonning ish unumdorligini sezilarli darajada oshirishi mumkin va u matndagi terminologiyaning bir xilligini saqlash ishini soddalashtiradi. Tarjima xotirasi tarjima qilishda samarali bo‘lib, ko‘p sonli takroriy matnlar (ma‘lum bir mavzu bo‘yicha statistik ma‘lumotlarga ko‘ra, tarjima uchun barcha materiallarga nisbatan bunday matnlarning ulushi 70% ga yetishi mumkin, bir xil hujjatning turli xil versiyalari) va mahalliyashtirish dasturlarida faol foydalaniladi. Tarjima drayveri kam sonli takroriy matnlarni tarjima qilishda samarasiz usul bo‘lib, shuningdek, foydalanuvchilarni o‘qitish uchun vaqt va moliyaviy xarajatlarni talab qiladi.

TM dasturlaridan foydalanish, shuningdek, har qanday axborot texnologiyalaridan foydalanish yangisini rad etish muammosidan kelib chiqadigan qiyinchiliklarga duch kelishi mumkin. Bu muammo zamonaviy rus jamiyatining mentaliteti sharoitida isbotlangan masalalardan

biridir. Shu bilan birga, bu omilni yengish zamon ruhi bilan bog'liq. Bunga misol qilib EnRus tarjima agentligi tomonidan olib borilgan tadqiqot natijalariga ko'ra, tarjimonga qo'yiladigan asosiy talablardan biri TM dasturlari mavjudligi va ulardan foydalanish qobiliyati ekanligini ko'rishimiz mumkin. Shu munosabat bilan tarjima xotirasi dasturlarida bilimlarni boshqarish usullari va usullarni o'quv jarayoniga integratsiyalashuvi hozirgi vaqtda kelajakdagi tarjimoning mehnat bozorida raqobatbardoshligi masalasidir. I.O.Babina fikriga ko'ra ta'limda tarjima xotirasi texnologiyasidan foydalanish o'quvchilarda amaliy ko'nikmalarni shakllantirish va ularni bozor talablariga tayyorlashda samarali vosita hisoblanadi [Бабина.О.И., 2013:98-108].

Parallel korpusdan ma'lumotlar bazasi sifatida jamlangan tarjima xotirasi muhitidan foydalanish tarjima jarayonini tezlashtiribgina qolmay, uning sifatiga ham ijobiy ta'sir ko'rsatadi. Zero matnda yo'l qo'yilgan xatolarni ko'rsatish va to'g'irlash, qolaversa, turli uslubga tegishli matnlardan ma'nodosh so'zlarning mosini qo'yishda ushbu dasturning imkoniyatlarini qo'llash mumkin.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Kompyuter lingvistikasi (darslik). – Toshkent: Nodirabegim; - T., 2021.
2. Abduraxmonova N. Mashina tarjimasining lingvistik ta'minoti. (Monografiya) Toshkent – 2018.
3. Abdurakhmonova N.Z., Shamsiyeva G.A. Theoretical foundations of corpus-based Uzbek-English machine translation // 2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE). – 2024. – <https://10.1109/PIERE62470.2024.10805010>
4. Abdurakhmonova N, G. Hayrullayeva, K.Urdishev Using A Corpus For Teaching Synonymy In Uzbek Электронная письменность народов российской федерации: опыт, проблемы и перспективы Материалы II Международной научной конференции (Уфа, 11–12 декабря 2019 г.) Russian, Ufa.
5. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
6. Бабина.О.И, Осминн.П.Г. Память переводов в обучении переводчиков. Вестник ЮУрГУ. Серия «Образование. Педагогические науки» 2013, том 5, № 3. - С.98-108. Qarang: <https://cyberleninka.ru/article/n/pamyat-perevodov-v-obuchenii-perevodchikov>
7. Bowker L. *Computer-Aided Translation Technology: A Practical Introduction*. University of Ottawa Press. 2002.
8. Shamsiyeva G. Linguistic basis of parallel corpus in the creation of MT models for Uzbek-English languages // So'z san'ati xalqaro jurnali. – 2024. – T. 7, №3. – B. 80–85. – <https://10.26739/2181-9297> – ISSN 2181-9297.

ВЫРАЖЕНИЕ ЯЗЫКОВОГО АТОМА «РОД» НА ГРАММАТИЧЕСКОМ УРОВНЕ УЗБЕКСКОГО ЯЗЫКА (когнитивно-вербальный аспект)

Йигиталиев У.С.

доктор философии по филологическим наукам,
доцент кафедры узбекского языка
Кокандского государственного университета

Умиджонов Ш.У.

магистрант Национального университета Узбекистана.

Annotatsiya: mazkur ilmiy maqolada o'zbek tilining grammatik sathida "jins" lingvistik atomining voqelanishi kognitiv-verbal aspektda tahlil etiladi.

Kalit so'zlar: lingvistik birlik, lingvistik atom, kognitiv-lingvistik aspect, grammatik sath, "jins" lingvistik atomi.

Аннотация: В данной научной статье анализируется возникновение языкового атома «род» на грамматическом уровне узбекского языка с когнитивно-вербального аспекта.

Ключевые слова: языковая единица, языковой атом, когнитивно-лингвистический аспект, грамматический уровень, языковой атом «род»

Abstract: This scientific article analyzes the emergence of the linguistic atom "gender" at the grammatical level of the Uzbek language from the cognitive-verbal aspect.

Keywords: linguistic unit, linguistic atom, cognitive-linguistic aspect, grammatical level, linguistic atom "gender"

Объединение лингвистических единиц в определённые парадигмы на основании определённого общего значения породило теорию поля.

Истоки ее зарождения приходятся на 20-30-е годы прошлого столетия, что связано с новым толкованием теории В.Гумбольдта «о внутренней форме языков». Научные дискуссии по вопросам о внутренней форме языков, являющихся основным объектом крупномасштабных лингвистических исследований того периода, стали основой создания подобной теории.

В последнее время проделана большая работа в области исследования проблем теории поля. По утверждению Г.С.Щура, опубликовано свыше тысячи статей, посвященных теории поля. В узбекском языкознании также появились работы, связанные с теорией поля. На наш взгляд, данный метод способствует более подробному изучению масштаба реализации лингвистических единиц. Рассмотрим понятие лингвистический компонент – «атом» (утверждение наше.—У.Й.), а также его семантико-функциональную реализацию в лингвистических ярусах узбекского языкознания. Лингвистический атом, как и другие лингвистические единицы (ЛЕ), имеет особенность разложения, обособленности для носителей определенного языка. Рассмотрим лингвистический атом (ЛА) *род*. Однако, прежде чем осветить масштабы семантико-функциональной реализации ЛА *род* в лингвистических ярусах, рассмотрим само понятие «*род*».

В узбекском языке слово «жинс» заимствовано из арабского. В «Толковом словаре узбекского языка» даются следующие его определения:

1) *биол.* – один из двух (самка и самец) видов живых организмов по биологическим и фразеологическим особенностям: *Эркак жинс. Хотин жинси*; 2) вид, сорт, разновидность;

3) социальная группа, сословие; род, происхождение, принимать в свою среду: *Ота, бу синглим, –табассум билан деди Йўлчи. – Бойникида яшаши маъқул кўрмадим. – Яхши қилибсан: Кабутар бо кабутар, жинс бо жинс. Тўзри улар қариндошинг, лекин улар сени ўз жинсига қўймайди. Улар пул жинсидан.* (Ойбек, «Кутлуг қон»). (*Верно, они твои родственники, но тебя они не примут в свою среду. Они из рода денег*); 4) геол. – минерал, образующий определённый слой земной коры. Горные породы.

Из общего числа определений в качестве объекта исследования мы выбрали лишь понятие «биологический род».

Лингвистический атом *рода* – это некое общее, специфически реализующееся в лексико-семантической системе нашего языка. Ранее [8: ст.55], нами было проанализировано это лексико-семантическое общее. В данной статье рассмотрим функциональную реализацию данного концепта в грамматической системе. Грамматическая система, как правило, рассматривается в качестве комплекса синтаксического и морфологического ярусов. Так как в предыдущей указанной статье нами были приведены масштабы семантико-функциональной реализации ЛА *род* в самостоятельных и служебных лексемах, здесь мы их не будем анализировать. Действительно, в грамматико-функциональном плане морфологических категорий не встречается лингвистический концепт *род*. Поэтому основное внимание мы уделили функциональной реализации синтаксического яруса. В синтаксической системе, особенно в простых предложениях, ЛК рода проявляется очень ярко. Например, ЛА мужского рода в семантике повествовательных предложений: *Салқин шамолларгина қизнинг сочлари билан ўйнашади. Юрагим сен эдинг, Гулнор* (*Прохладный ветерок играет волосами девчонки. Ты была моим сердцем, Гульнар*); в семантике вопросительных предложений: *Оппоқ қиз, қаерга борамиз?, Нима маслаҳат берасиз, Ойгул?* (*Девушка, куда пойдём? Что скажешь, Айгуль?*), в семантике побудительных предложений: *Э, халойиқ, бутун кучимизни бир куч қилиб тўплайлик! Куннинг тигида нима қиласан, салқинда ўтирсанг-чи!* (*Друзья, соберем свои силы в единый кулак! Зачем ходить в такую жару, сиди дома!*), либо ЛА женского рода в составе повествовательных предложений типа. *Вой, овсин-эй, хомсиз-да ҳали. Эринг гапига лаққа тушиб юраверасиз* (*Эх, вы, подружка, ещё не вы не опытни. Верите каждому слову мужа.*) и вопросительно-побудительных предложений типа *Нигорани, яхши, ўзига тўқ хонадонга берамиз, дадаси?! (Нигору отдадим замуж за хорошего, обеспеченного человека).* ЛА *род* может повторяться также в грамматико-функциональном плане сложных предложений и крупных синтаксических единиц – *текстов*. Например:

1. *Вой, ўлай, шунча ёшга кириб, совчиликка эркак киши келганини биринчи кўришим* [С.Кароматов, романа «Олтин кум». –С.174.] (*Боже мой, столько прожив, впервые слышу, чтобы сватать приходил мужчина*).

2. *У тўй қилсам бировдан сўрарлик бўлмасин, деб чинни идишларни ҳам етарли қилиб қўйган эди*[Саид Аҳмад, роман «Уфк». – С.123]. (*Она достаточно запаслась фарфоровой посудой, чтобы в случае свадьбы не спрашивать у других*).

3. *У жон ҳолатда, қандай ўтирган бўлса шундай – яланг оёқ, яланг бош, уйдан югуриб чиқди*[Одил Ёкубов, роман «Диёнат». –С.151]. (*Она в чём была – без платка и без тапочек выбежала из дома*). В функциональном плане данных синтаксических конструкций скрыто понятие женского рода. А в следующих сложных предложениях ведущую позицию занимает ЛК мужского рода. Например:

1. – *Эҳтиёт бўлинг, хоним, – совуққонлик билан гапирди йигит, жонингизга жабр қилиб қўясиз* [А.Дюма, роман «Уч мушкетёр». – С.97.] (*Будьте осторожны, мадам, –сказал парень, –вы можете навредить себе*).

2. *Хотин, юравер, Тогага айтдим, майли деди* [Саид Аҳмад, роман «Уфк». – С.99] (*Идём, жена, я сказал дяде, он согласен*).

Грамматико-функциональный аспект следующих текстов подтверждает, что целый текст содержит ЛА женского или мужского рода:

1. Ҳа... Э, бу кишини тирик тарих деса бўлади! Шу Тошкентда, Бешёғоч даҳасидаги машҳур Маъдихон эшон мадрасасида таҳсил кўрганлар. Кейин оқ пошшо-герман уриши бошланиб, қишлоқда дўкандорлик қилган оталари «синганлар» у бу киши бир кундаёқ, мулламиркуруқ муллаваччага айланиб, мардикорликка кетганлар [Одил Ёкубов, роман «Диёнат». – С.142] (Да... Этот человек – живая история! Он получил образование здесь, в Ташкенте. В известном медресе Маъдихан на Бешагаче. Затем началась русско-германская война, его отец, который имел небольшую лавку в кишлаке, разорился, а этот человек был вынужден стать подёзником).

2. У ёшлигида текканга тегиб, тегмаганга кесак отарди. Атлас кўйлагини қора мойга белаб тракторчининг орқасида юрар, у тушликка чиққанда сўрамай тракторни ҳайдаб қоларди. Қишлоқ кўчаларини чангитиб, товуқларни қақалатиб яна жойига келтириб қўйиб, ғойиб бўларди [Одил Ёкубов, роман «Диёнат». – С.171.] (В молодости она никому спуска не давала. В запачканном масле атласном платье она ходила за трактористом, а когда он уходил на обед, без спросу садилась на трактор и угоняла его. Поднимая пыл на улицах кишлака, распугивая кур, возвращала трактор на место и бесследно исчезала).

Таким образом, приведённые примеры доказывают верность нашего научного предположения о выражении ЛА *род* в данном ярусе и образовании на основе этого особых концептуальных групп.

Иными словами, это гармонирует с процессом превращения в лингвистическую истину естественнотехнического явления «валентности» 40-х годов прошлого столетия.

Список использованной литературы:

1. Гумбольдт В. фон. Избранные труды по языкознанию: Пер. с нем. – М.: Прогресс, 1984. – 397 с.
2. Щерба Л.В. Языковая система и речевая деятельность. –Л.: Наука, 1974. – 451 с.
3. Yigitaliyev U.S. O‘zbek tilida “odam” va “jins” konseptlari bilan bog‘liq assotsiativ-verbal birliklar: Filol. fan.bo‘y. fals. dokt. ...(PhD) diss. avtoref. – Toshkent, 2020. – 54b.
4. Yigitaliyev U.S. O‘zbek tilida konseptning assotsiativ-verbal tadqiqi (Monografiya). – Toshkent: “Lesson press”, 2023. – 130 b.
5. Yigitaliyev U.S. The transposition of the concept of "man" in the uzbek language in the semantics of verb tense forms. ASEAN Journal on Science & Technology for Development Vol 39, No 4, 2022, 423-431 423.
6. Yigitaliyev U.S., Shohruhbek Umidjanov. The role of the concept "man" in the functional-semantic and assotciative-verb expression of the forms of the category of cases (on the example of Uzbek-language sources) // Journal for Educators, Teachers and Trainers, Vol. 14 (5) 18-22 / <https://10.47750/jett.2023.14.05.004>
7. <https://scienceweb.uz/publication/11972>
8. <https://docplayer.ru/27467248-Russkiy-associativnyy-slovar.html>;
9. <https://www.oarepo.org/index.php/oa/article/download/1428/1420>
10. <https://www.oarepo.org/index.php/oa/article/view/1320>
11. <https://scienceweb.uz/publication/12062>

PRAGMALINGVISTIKA VA UNDA DISKURS TUSHUNCHASI

Abdunazarova Maftuna Faxriddinovna

O'zMU 10.00.11 – Til nazariyasi.

Amaliy va kompyuter lingvistikasi tayanch doktranti

E-mail: abdunazarova_m@nuu.uz

Annotatsiya: Ushbu maqolada pragmalingvistika fanining shakllanishi, asosiy tamoyillari hamda diskurs tushunchasining mazmun-mohiyati, uning pragmalingvistik yondashuvdagi o'рни tahlil qilinadi. Turli nazariy qarashlar asosida diskursning ko'p qirrali tabiati yoritiladi va tilshunoslikda uni o'rganishning dolzarbligi asoslab beriladi.

Diskurs tahlilining asosiy yondashuvlari va metodlari ko'rib chiqiladi. Turli mualliflar tomonidan ilgari surilgan nazariyalar solishtirilib, diskursga zamonaviy yondashuvlar asosida baho beriladi. Nutqiy aktlar, kontekst, presuppozitsiya va ishoraviylik kabi jihatlar misollar orqali ochib beriladi.

Kalit so'zlar: Pragmalingvistika, diskurs, diskurs tahlili, kontekst, nutq, kommunikatsiya, dialogik diskurs, nutq akti, presuppozitsiya.

Abstract: This article examines the formation of pragmalinguistics, its main principles, and the concept of discourse within this framework. It highlights the multifaceted nature of discourse and its relevance in linguistic research, based on various theoretical perspectives.

The main approaches and methods of discourse analysis are considered. The theories put forward by different authors are compared, and the discourse is evaluated based on modern approaches. Such aspects as speech acts, context, presupposition, and allusion are revealed through examples.

Keywords: Pragmalinguistics, discourse, discourse analysis, context, speech, communication, dialogic discourse, speech act, presupposition.

Аннотация: В данной статье анализируется формирование прагмалингвистики, ее основные принципы, сущность понятия дискурса и его роль в прагмалингвистическом подходе. На основе различных теоретических взглядов освещается многогранная природа дискурса и обосновывается актуальность его изучения в лингвистике.

Рассматриваются основные подходы и методы анализа дискурса. Сравняются теории, выдвинутые разными авторами, и дискурс оценивается на основе современных подходов. Такие аспекты, как речевые акты, контекст, presupпозиция и жесты, раскрываются на примерах.

Ключевые слова: Прагмалингвистика, дискурс, анализ дискурса, контекст, речь, коммуникация, диалогический дискурс, речевой акт, presupпозиция.

Zamonaviy tilshunoslikda insonning til orqali muloqotga kirishish jarayonlarini o'rganishga alohida e'tibor qaratilmoqda. Ayniqsa, tilning qo'llanilishi, ya'ni nutqiy faoliyatni tahlil qilishda pragmalingvistika muhim o'rin tutadi. Pragmalingvistika, asosan, nutq ishtirokchilarining maqsadi, nutqiy akti, niyati, kommunikativ vazifalari kabi omillarni hisobga olgan holda til birliklarini tahlil qiladi. Shu bilan birga, "diskurs" tushunchasi ham bu sohada markaziy o'ringa ega bo'lib, muloqot jarayonining kontekstual asosda tahlil qilinishini ta'minlaydi.

XX asrning 50-yillaridan so'ng tilshunoslikka pragmatik yondashish keng tus oldi. Pragmatika – pragmatik lingvistik, pragmalingvistika, pragmafonologiya, pragmafonetika, pragmagrammatika, pragmasintaksis kabi yangi tushuncha va atamalar, bu tushuncha va atamalar ostida esa til hodisalariga yangicha yondashish, ularni yangicha talqin qilish bilan bog'liq bo'lgan yangi tushunchalar va tahlil usullari keng ommalashdi.

Pragmatikaning lugʻaviy maʼnosi yunoncha “pragma” soʻzidan tarjima qilinganda ish, harakat maʼnolarini anglatadi. “Pragmatika” termini XX asr 30-yillarining oxirida Ch. U. Morris tomonidan semiotikaning bir boʻlimi sifatida fanga kiritilgan. Ch. Morris semiotikani uchga boʻlgan: 1. Semantika – belgilarning obyektga boʻlgan munosabatini oʻrganadi. 2. Sintaksis – belgilararo munosabat. 3. Pragmatika – soʻzlovchi til belgilariga munosabatini tadqiq qiluvchi boʻlim. Dastlabki izohlarda ushbu uch soha yonma-yon joy olgan boʻlsa-da, Morris keyinchalik “pragmatika” tushunchasini qolgan ikki soha - sintaksis va semantikaga nisbatan ancha keng ekanligini qayd etadi.

Baʼzi manbalarda esa pragmatika falsafiy tushuncha sifatida fanga kirib kelganligi aytib oʻtiladi. Jumladan, “pragmatika aslida falsafiy tushuncha boʻlib, u Sokratdan oldingi davrlarda ham qoʻllanishda boʻlgan va keyinchalik uni J. Lokk, E. Kant kabi faylasuflar Aristoteldan oʻzlashtirganlar. Shu tariqa falsafada pragmatizm oqimi yuzaga kelgan. Bu oqimning asosiy taraqqiyot davri XIX - XX asrlardir. Ayniqsa, XX asrning 20-30- yillarida pragmatizm gʻoyalarining keng targʻibi aniq sezila boshladi. Amerika va Yevropada bu targʻibotning keng yoyilishida Ch. Pirs, R. Karnap, Ch. Morris, L. Vitgenshteyn kabilarning xizmatlarini alohida qayd etish kerak¹¹.

Pragmatikaning tilshunoslikka munosabati haqida fanda xilma-xil qarashlar mavjud. Ayrim tilshunoslar, chunonchi, V. Dresslar “Pragmatika tilshunoslikka aloqador emas” deb yozadi¹². Lekin pragmatikani tilshunoslikdan batamom uza-digan olimlar sanoqli. Ularning koʻpchiligi pragmatikaga tilshunoslikning bir oqimi va nutq jarayonlarining alohida bir qirrasini oʻrganishga va tavsif etishga ixtisoslashgan yoʻnalishi sifatida talqin qilishadi.

Pragmatika tilshunoslikning bir tarmogʻi, aniqroq qilib aytganda, nutq birliklarining nutq jarayonida boshqa nolisoniy vositalar bilan birgalikda qoʻllanishini, uning samaradorligini oʻrganuvchi fan tarmogʻidir.

Muloqot jarayonida soʻzlovchilar tomonidan tilning qoʻllanilishi va soʻzlovchilarning pragmatik bilimlarini birlashtiruvchi hodisa sifatida pragmatika ritorika, stilistika, nutq nazariyasi va tipologiyasi, nutq faoliyati, sotsiolingvistika, psixolingvistika va shu kabi doiralarda uzoq oʻrganilish tarixiga ega boʻlgan koʻplab muammolarni qamrab oldi.

Pragmatika nutq va axloqiy, insoniy xatti-harakat, ruhiy faoliyat bilan bogʻliq boʻlgan koʻplab fanlarning kesishish nuqtasidadir. Pragmatika struktur tilshunoslik, stilistika, nutq madaniyati, poetika va lingvopoetika, psixolingvistika, etika va estetika kabi fanlar bilan, umuman insonning ongli va ongsiz madaniy va tabiiy xatti-harakati bilan bogʻliq boʻlgan, kulturologiya umumiy nomi bilan birlashadigan fanlarning barchasi bilan yaqin aloqada boʻladi. Ammo ularning birortasi pragmatikaning oʻrnini toʻla-toʻkis bosa olmaydi.

Pragmatikaning “qayta tugʻilish davri” deb eʼtirof etilgan 1970-yildan boshlab, xorijiy tilshunoslikda haqiqiy pragmatik koʻtarilish yuzaga keldi. Bu mavzuga oid qator anjumanlar, yigʻinlar oʻtkazildi. Bu darajada jadal taraqqiy etayotgan fan sohasining toʻligʻicha shakllanishi, oʻzligini ajratib olishi uchun uning asosiy tamoyillari, tushunchalari, predmetini belgilab olish zarurati yuzaga keldi. Jumladan, diskurs tushunchasi va uning turlari ham pragmalingvistikaning asosiy predmetlaridan biri sifatida tadqiq etib kelinmoqda.

Diskurs tushunchasi: mazmuni va talqinlari

Diskurs bugungi kunda lingvistika, sotsiologiya, psixologiya, falsafa kabi fanlararo tadqiqotlarda keng qoʻllanilayotgan tushuncha boʻlib, u nafaqat tilshunoslikda, balki ijtimoiy munosabatlar va ongni ifodalash vositasi sifatida ham muhim ahamiyatga ega.

Diskurs termini ilk bor fransuz faylasufi Mishel Fuko tomonidan keng muomalaga kiritilgan. Keyinchalik tilshunoslikda T.A. van Deyk, Norman Ferklof, Ruth Vodak, J.R. Searle kabi olimlar diskursga metodologik asos yaratishgan.

Diskurs — bu faqat matn emas, balki muloqot davomida yuzaga keladigan, kontekst bilan boyitilgan til hodisasi sifatida qaraladi. Diskurs nafaqat lingvistik, balki ijtimoiy, madaniy, siyosiy

¹¹ Safarov Sh. Pragmalingvistika. – Toshkent, 2008. – B. 40-41.

¹² Safarov Sh. Nutqning etnosotsiopragsmatik tahlili asoslari. – Samarqand, 2007. – B. 18-19.

va boshqa omillarni ham o'z ichiga oladi. Bu tushuncha orqali til hodisalariga ko'proq ijtimoiy-kommunikativ hodisa sifatida yondashiladi. Shu sababli diskursga oid talqinlar ko'p va xilma-xildir: Fuko, van Deyk, Fairclough kabi olimlar bu tushunchaga turli nazariy yondashuvlar ishlab chiqqanlar.

Diskurs tahlilining asosiy yo'nalishlariga kognitiv, kritikal, sotsiopragmatik va funksional yondashuvlarni kiritish mumkin. Masalan, kritikal diskurs tahlili til orqali ijtimoiy tengsizlik va hokimiyat munosabatlarni ochib berishga qaratilgan.

Pragmalingvistik yondashuvda diskurs tahlili nutq ishtirokchilari, ularning niyatlari, rollari, kontekst va ijtimoiy vaziyatlarni hisobga olgan holda amalga oshiriladi. Masalan, bir matnda ishlatilgan gapning ma'nosi faqat so'zlar orqali emas, balki aytilgan vaziyat, kim aytganligi, qachon va qanday sharoitda aytilganligiga qarab aniqlanadi. Bu esa diskursni tahlil qilishda an'anaviy grammatik yondashuvdan farqli ravishda, kengroq kommunikativ ko'nikmalarni talab etadi.

Diskurs doimo muayyan kontekstda vujudga keladi. Kontekst — bu muloqot ishtirokchilari, ularning ijtimoiy maqomi, nutq vaziyati, madaniy muhit kabi omillarni o'z ichiga oladi. Diskursning interaktiv xususiyati esa uning dinamik va o'zgaruvchanligini ta'minlaydi.

Hozirgi vaqtda diskurs tahlili faqat adabiy yoki siyosiy matnlar emas, balki internet muloqoti, ijtimoiy tarmoqlardagi postlar, ommaviy axborot vositalari tili kabi sohalarda ham qo'llanilmoqda. Bu jarayonda kompyuter lingvistikasi, korpus lingvistikasi kabi yondashuvlar ham diskurs tahliliga jalb etilmoqda. Shu bois zamonaviy pragmalingvistika tahliliy, kognitiv, funksional yondashuvlar bilan birgalikda rivojlanmoqda.

Diskurs tahlilining metodologiyasi

Diskurs tahlili pragmalingvistika, sotsiolingvistika, semiotika, psixolingvistika kabi sohalar bilan bevosita bog'liq. Asosiy metodlar quyidagilardan iborat:

- Kontekstual tahlil
- Nutq strategiyalarini o'rganish
- Nutqiy akt tahlili
- Ijtimoiy rol va statusning ta'siri

Diskurs odatda quyidagi turlarga bo'linadi:

- Dialogik diskurs
- Monologik diskurs
- Ilmiy diskurs
- Ommaviy axborot diskursi
- Ijtimoiy-siyosiy diskurs

Har bir diskurs turi o'ziga xos stilistik, funksional va kommunikativ xususiyatlarga ega.

Dialogik diskurs zamonaviy tilshunoslikning eng faol o'rganilayotgan yo'nalishlaridan biridir. U til va jamiyat o'rtasidagi o'zaro bog'liqlikni ochib berishda muhim o'rin tutadi. Dialog — bu ikki yoki undan ortiq ishtirokchi o'rtasidagi fikr almashinuvi bo'lib, u o'z-o'zidan diskursning interaktiv va kontekstual xususiyatlarini namoyon etadi.

M. Baxatin dialogiklikni inson tafakkurining asosiy shakli sifatida talqin qilgan. Dialogik diskurs bu — ikki tomonlama muloqot, ya'ni bir kishining nutqi boshqasining so'ziga javob bo'lib xizmat qiladi.

Diskurs tahlili — bu tilshunoslikning zamonaviy yo'nalishlaridan biri bo'lib, til va jamiyat o'rtasidagi murakkab munosabatlarni o'rganishga xizmat qiladi. U muloqot jarayonining nafaqat lingvistik, balki ijtimoiy, madaniy va kontekstual jihatlarini chuqur tahlil qilish imkonini beradi.

Diskurs tahlili nafaqat matn va nutqning til birliklari darajasida balki ularning ortidagi ma'no va niyatlarni anglash imkonini beradi. Bu esa tilshunoslikni yanada chuqurlashtirib, uni ijtimoiy fanlar bilan integratsiyalashga yo'l ochadi.

Pragmalingvistika va diskurs tushunchasi tilshunoslikdagi kommunikativ yondashuvlar asosini tashkil etadi. Ular orqali til hodisalarini yanada chuqurroq va real muloqotga mos holda

tahlil qilish mumkin bo‘ladi. Maqolada ko‘rib chiqilgan nazariyalar va amaliy yondashuvlar pragmlingvistik tahlilning zamonaviy tilshunoslikdagi ahamiyatini ochib beradi.

Foydalanilgan adabiyotlar

1. Abdurakhmonova, N. Z., & Teshabayev, A. G. (2019). Modeling IT of discourse analysis and the issues machine translation. Scientific Bulletin. Physical and Mathematical Research, 1(2), 101-107.
2. Morris, Ch. Foundations of the Theory of Signs. – University of Chicago Press, 1938.
3. Lakoff G. Pragmatics in natural logic. – Formal semantics of natural language. – Cambr., 1975, – P. 171.
4. Yule.G. Pragmatics. – Oxford. Oxford Univ. Press, 1996. – P.138.
5. Safarov Sh. Pragmlingvistika. – Toshkent, 2008. –B.125.
6. Safarov. Sh Nutqning etnosotsiopragsmatik tahlili asoslari. – Samarqand, 2007. – B.18-19.
7. Hakimov M. O‘zbek pragmlingvistikasi asoslari. – Toshkent.: Akademnashr, 2013. – B. 120.
8. Van Dijk, T.A. Discourse as Structure and Process. 1997.
9. Fairclough, N. Critical Discourse Analysis. 1995.

“QIYOMAT” ROMANIDAGI FRAZEMALARNI PARALLEL KORPUSDA LINGVISTIK TEGGLASH MASALASI

Iskandarova Aybibi

O‘zbekiston Milliy universiteti dotsenti, f.f.n.

E-mail: aybibiiskandarova1962@gmail.com

Annotatsiya: Ushbu maqolada parallel korpusda rus-o‘zbek frazemalarining xoslanishi, tarjima jarayonida interpretatsiya holatlari, ularning birikma va gap holatidagi shakllari, lingvistik teglash masalalari haqida mulohaza yuritiladi.

Kalit so‘zlar: parallel korpus, frazema, kolokatsiya, xoslanish, so‘z birikmasi va gap shaklidagi frazema, morfologik modellar, teglash.

Аннотация: В статье рассматриваются особенности русско-узбекских фразеологических единиц в параллельном корпусе, ситуации интерпретации в процессе перевода, их формы в ситуациях сложноподчинения и предложения, а также вопросы языковой маркировки.

Ключевые слова: параллельный корпус, фраза, коллокация, атрибуция, фраза в форме словосочетания и предложения, морфологические модели, разметка.

Annotation: This article discusses the localization of Russian-Uzbek phrasemes in a parallel corpus, interpretation cases in the translation process, their forms in compound and sentence forms, and issues of linguistic labeling.

Keywords: parallel corpus, phraseme, collocation, localization, phraseme in the form of a word combination and sentence, morphological models, labeling.

O‘zbek adabiy tilimizni boyitib borish va asrash, uni avlodlarga sof holatda meros qoldirish, xalqaro nufuzini balandlatish har birimizning burchimizdir. O‘zbek parallel korpusini bir necha tillar kesimida turli janrga doir matnlar bilan boyitib borish, shungdek, parallel frazemalarning morfologik va sintaktik modellashirilgan, tarkibiy xususiyatlariga ko‘ra teglangan ta‘minotini yaratish, ularning lingvokulturologik asoslarini aniqlash va rus-o‘zbek til egalari lisoniy ongi bilan bog‘liq kognitiv modellarni ishlab chiqish, mazkur tillardagi frazemalarni qiyoslash jarayonida olingan xulosalardan o‘zbek frazeologiyasi taraqqiyotining umumiy va xususiy belgilarini aniqlash soha mutaxassislarining vazifasidir.

Xalq tafakkuri, uning olamni idrok etishidagi o‘ziga xos jihatlari barqaror birikmalarda namoyon bo‘ladi. Bu lingvomadaniy birliklarda o‘sha xalq haqida madaniy axborot beriladi. V.A.Maslova frazeologizmlarning aksariyatida aniqlanishi lozim bo‘lgan milliy madaniyatning “izi” mavjudligi; frazeologik birliklarning ichki shaklida obyektiv olamning u yoki bu hodisalarini obrazli tasavvuri bo‘lgan madaniy informatsiya saqlanishi va u frazeologizmga madaniy-milliy bo‘yoq berishi; ularning tahlili jarayonida milliy-madaniy konnotatsiyani ochish muhimligini ta‘kidlaydi [1.82]. V.N. Teliya frazemalarni “lingvomadaniy jamoa milliy o‘zligini ko‘radigan ko‘zgudir” deb baholaydi [2.76].

Sh. Rahmatullayev frazeologik birliklarning tarkibida kamida ikkita leksema qatnashishi; o‘zaro sintaktik bog‘langan bu leksemalar mohiyatiga ko‘ra birikmaga yoki gapga teng bo‘lishi; frazeologik ma‘no, uni sintaktik birlik (nutq birligi) deb emas, balki semantik birlik (til birligi) deb qarashga olib kelishini asoslaydi: “Iboralarning asosiy qismini fe‘l frazeologik birliklar tashkil etadi. Frazeologik birliklarning leksik tarkibida, fe‘l so‘z - komponentidan tashqari, boshqa turkum so‘zi ham qatnashadi. Bunday komponentlar ko‘pincha ot, sifat, ravish turkumi bilan ham ifodalanadi.” [3.9].

Professor N.Abduraxmonova mashina tarjimasi uchun ingliz-o‘zbek frazemalari modellarini tadqiq qilgan ekan, ingliz tilida frazema terminining quyidagi ma‘nolarda qo‘llanilishini

qayd etadi:1) fe'l yoki ravish bilan keluvchi predloglarning birgalikda muayyan ma'noga tenglashishi; predlog bilan keluvchi ot, sifat, fe'l va ravish so'z turkumlarining o'zaro yaxlitlanib, bitta ma'noda kelishi. Shuningdek, frazeologik birliklarning frazema va idiomalarga ajratilishi, idiomalar tarkib toptiruvchi komponentlarining yaxlitlanib, bir umumiy ma'noni ifodalashi, turg'un birikmalarining idioma, frazeologizm, frazema, frazeologik birliklar nomlari bilan yuritilishini ta'kidlaydi [4. 36].

Frazeologizm, frazeologik birlik, frazema – ikki yoki undan ortiq so'zdan tashkil topgan, ma'noviy jihatdan o'zaro bog'liq, so'z birikmasi yoki gapga teng keladigan, yaxlitligicha ko'chma ma'noda qo'llanadigan barqaror (turg'un) bog'lanmalarining nomidir. Tilshunoslikda frazeologizmlar ibora, frazeologik birikma, turg'un birikma, barqaror birikma atamallari bilan yuritiladi. Barqaror birikmalarga maqol va matallar ham kiradi.

Uzbekcorpus.uzda va rus, ingliz korpuslarida iboralar **Collocation /COLC/** atamasi bilan yuritiladi. Chingiz Aytmatovning “Qiyomat” romanida talaygina turg'un birikmalar mavjud bo'lib, ular birikma va gap shaklida qo'llanilgan. “Qiyomat” da turg'un birikmalarni tahlil qilganimizda , ularning tarkibi [ot+fe'l], [sifat+fe'l], [ot+sifat], [ot+ot] kabi shakllaridagi namunalardan iborat ekanligi ma'lum bo'ldi. Masalan, *милость судьбы* [ot+ot] – **taqdirning mehribonchiligi** [ot+ot] ; *проклятой травы* [sifatdosh+ot] – **qarg'ish tekkan bang o't** [sifatdosh +ot], *самый покладистый и безобидный* [sifat+sifat]– **qo'y og'zidan cho'p olmagan** [ot+ot+ot+fe'l].

Bir-biridan uzoq bo'lgan tillarga tegishli badiiy matnlar tarjimasida, bunday farqli jihatlar sodir bo'lishi tabiiy: *целая система промыслах* [sifat+ot+ot] – **miridan-sirigacha ishlab chiqmoq** [ravish+fe'l+fe'l]; Tarjima jarayonida ko'chma ma'noga ega bo'lmagan birliklar ibora shaklida o'girilishi mumkin: masalan, *как сумашедший* – **xuddi jin tekkandek, поменьше общаться между собой** – **miq etib og'iz ochmaslik, задуманное дело** – **ko'nglimga tukkan ish, он умер** – **omonatini topshirdi, обо всех своих неудачах** – **ishi yurishib ketmayotganligi, уже к вечеру** – **oqshomga tortib, не выдержали** – **yuraklari dosh bermay...**

Asarda frazemalarning eng ko'p uchraydigan shakli [ot+fe'l] bo'lib, korpus tilida fe'lli kolokatsiya deb yuritiladi, ularni quyidagi jadvalda ko'rsatamiz:

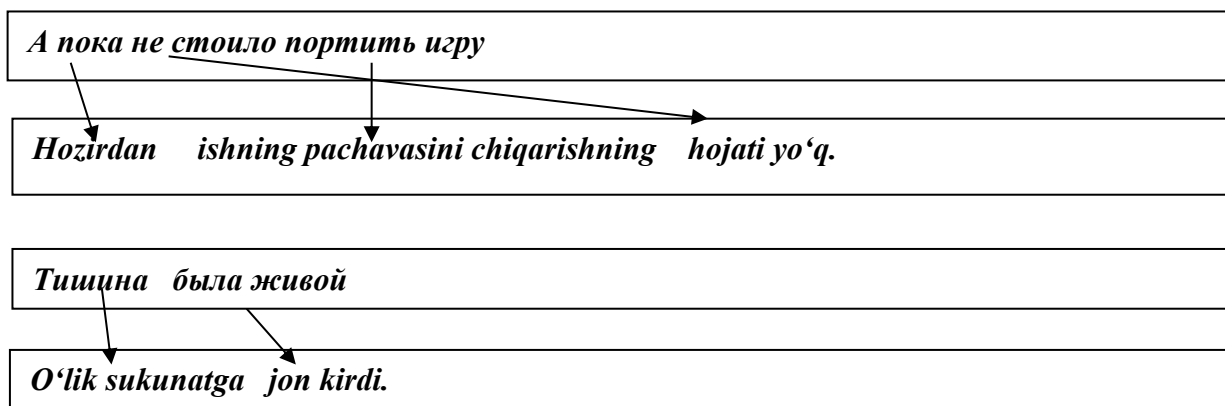
1	<i>притаилась, от страха</i>	esxonalari chiqqancha
2	<i>в страхе припустили еще сильнее</i>	vahimaning kuchi bilan yanada zo'r berib
3	<i>исхлестанным ветром лицом</i>	basharasini yel yalab ketgan
4	<i>и понесли куда глаза глядят</i>	bosh oqqan tomonga jo'namoq
5	<i>озлоблены на мир</i>	dunyodan to'yg'an
6	<i>не разгадали при этом извечной загадке</i>	boshlarini qotirib kelayotgan jumboqni yecha olmayotganliklari
7	<i>заново восстанавливая в памяти</i>	xotirasida yana bir boshdan tiklab
8	<i>замысел состоял в том</i>	ko'ngliga tugib qo'yomoq
9	<i>надеялся на память</i>	xotiramga ishonmoq
10	<i>не вызывают энтузиазма</i>	ko'ngillari sovimoq
11	<i>под свою ответственность</i>	mas'uliyatni o'z ustimga olib
12	<i>поднять в прессе</i>	matbuotda ko'tarib chiqmoq
13	<i>от саморазрушения личности</i>	shaxsning o'z-o'zini yemirishidan tortib
14	<i>ни за что не выдавать друг друга</i>	bir-birlarini sotmaslik
15	<i>как щепку, живой водоворот</i>	girdobga tushgan cho'pdek chirpirak bo'lib aylanmoq
16	<i>как между жерновами</i>	tegirmon tosh o'rtasida qolganday

Tarjimashunos olim G'. Salomov har bir ijodkorning o'ziga xos individual uslubidan kelib chiqadigan, “tarkibiy milliylik”, “uslubiy milliylik”, “individual milliylik”, “davriy milliylik”,

“nutqiy milliylik” mavjudligini, buni tarjimada qanday aks ettirish yo‘llarini izlab topish faqat tashqi milliy-etnografik belgilar, milliylik komponentlarini berish yo‘llarini aniqlashga qaraganda beqiyos murakkab va muhim” ligini alohida ta’kidlab o‘tadilar [5.83]. Asliyat tilidan maqsad tiliga tarjima jarayonida tarjimon qaysi xalqqa mansub bo‘lsa, o‘sha xalqning dunyoqarashi, fikrlashi, e’tiqodi akslanadigan milliy til qoliplaridan foydalaniladi: Chingiz Aytmatov “Плаха”da nashavand bolalarga nisbatan *мальчишек* so‘zini ishlatgan, tarjimon uni *ona suti o‘g‘zidan ketmagan* iborasi shaklida o‘g‘irgan, nashavand bola Pashaning otasiga nisbatan ishlatilgan *человек резкого* birikmasi *qattiqqo‘l*; *та трагедия - qora kun, xunbiyron qirg‘in solgan bir kuch* tarzida o‘g‘irilgan. Shuningdek, tarjimon romanda modal munosabatlarni ifodalaydigan frazemalardan o‘rinli foydalangan: *как бы то ни было – qanday bo‘lmasin; а он, несчастный – boyaqish yigit; но это к слову – bu otadi bir gap-da; моему наивному возмущению – voy bolasi tushmagur-eu; что другого исхода, к сожалению, нет – bo‘lar ish bo‘libdi, endi iloj qancha; Божье – Oh, Xudoyim.*

Badiiy tarjima interpretatsiya jarayonidir, ya’ni tarjimonning estetik e’tiqodi tarjimada iz qoldiradi: “Tarjima davomida tarjimon hamma vaqt o‘z ona tilida fikrlaydi, ona tili unga tahlil quroli, sinov mezon bo‘lib xizmat qiladi. Ayni vaqtda xuddi shu ona tili ijodiy sintez qilishga yordam beradi”[5. 98]. Quyidagi misollarda tarjimonning “uslubiy milliyligi”, “individual milliyligi,” tarjima birliklarini nechog‘lik o‘z o‘rnida mohirona ishlata olganligida ko‘rinadi:

Тишина была живой frazemasi - O‘lik sukunatga jon kirdi tarzida semantik va tarkib jihatdan mos holatda tarjima qilingan. *Прежде всего это были люди бездомные, перекаму – поле...* barqaror birikmasi - *Bedananing uyi yo‘q, qayga borsa, pitpildiq deganday, bular ham tayini yo‘q, qamg‘oqdek shamol qayerga uchirsa, shu yerda daydiydigan kimsalar edi* – tarkib jihatdan mos bo‘lmagan, obrazli,xalq tilida ko‘p ishlatiladigan ekvivalentdan foydalangan holda tarjima qilingan.



Frazemalar murakkab struktur-grammatik xususiyatga ega bo‘lgan yaxlit, yangi lug‘aviy ma’noni ifodalaydigan til birligidir. Bu ma’noni anglash esa avtomatik tarjima jarayoni uchun qiyin kechadi. Soha mutaxassislari frazemalarni maxsus teglar, razmetkalar bilan belgilanishi lozimligini ta’kidlaydilar: “Badiiy asarlar asosidagi parallel matnlar korpusida frazemalar bilan bog‘liq muammolar bo‘lishi aniq. Frazemalar ikki yoki undan ortiq so‘zdan tuzilib, shu so‘zlar semantikasi asosida yangi ko‘chma lug‘aviy ma’noni shakllantirishga xizmat qiladi. Shu bois, parallel matnlar korpusini shakllantirishda to‘g‘ridan to‘g‘ri mashina tarjimasidan foydalanilsa, semantikasi to‘g‘ri yoritishda muammolar kelib chiqadi. Muammolarni barataraf etish uchun frazemalar sodd va qo‘shma leksik birliklar, so‘z birikmalaridan farqlanishi, maxsus teglar, razmetkalar bilan belgilanishi lozim. Frazemalar tarjimasida, albatta, mutaxassis nazorati kerak bo‘ladi”[6.33].

Прежде всего это были люди бездомные, перекаати – поле, кроме, разумеется, Кепы: у троих из них ушли жены, все они были в той или иной степени неудачниками, а следовательно, были по большей части озлоблены на мир

Bedananing uyi yo‘q, qayga borsa, pitpildi deganday, bular ham tayini yo‘q, qamg‘oqdek shamol qayerga uchirsa, shu yerda daydiydigan kimsalar edi. Faqat Keparagina bundan mustasno edi. Uchtasini xotini tashlab ketgan, hammalari ham na bironvi yolchitgan, na o‘zlari yolchigan, shuning uchun ham dunyodan to‘ygan, hamma narsadan norozi

Rus-o‘zbek parallel korpusi doimiy ravishda kontentlar bilan boyitib boriladigan jarayonlardan iborat. Rus-o‘zbek tillaridagi frazemalarning tizimlashtirilgan va teglangan ma’lumotlar bazasi rus-o‘zbek tillarini o‘qitish, avtomatik tarjima jarayonini takomillashtirish, badiiy matnlar tarjimasini sifatli amalga oshirishda qo‘l keladi. Rus-o‘zbek tillaridagi frazemalar rus-o‘zbek parallel korpusining tizimlashtirilgan va teglangan ma’lumotlar bazasi ta’lim korpuslari, rus-o‘zbek tillarining bir necha tillar kesimidagi parallel korpuslarini yaratish uchun material bo‘lib xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Maslova V.A. Лингвокультурология: Учеб.пособия для студ. высш. учеб. заведений. - М. Издательский центр “Академия”, 2001. – С.82
2. Образные средства русского языка. Под. ред. В. Н.Телия. – М.Оточество. 1995. – 366 с.
3. Ш. Рахматуллаев. Ўзбек тилининг фразеологик луғати. Тошкент. 1992. 378 б.
4. Абдурахмонова Н. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (содда гаплар мисолида). Филол.фан.б. фалс. док. (PhD) дисс. – Тошкент, 2018. – Б.36.
5. Ғ.Саломов. Таржима назариясига кириш. – Тошкент, 1978. В.219.
6. Рахмонова А. Ўзбек тили миллий корпусини яратишда компьютер усуллари. Филол. фан.б. фалс.док. (PhD) дисс. - Тошкент, 2022. - Б.17.
7. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
8. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.

SEMANTIC-STRUCTURAL ANALYSIS OF EXPRESSIONS OF THE BEGINNING OF THE XX CENTURY

Kasimova Ziyoda Movlon qizi

PhD, National University of Uzbekistan

Abstract: This article analyzes phraseologisms with specific semantics and structure in linguistics. Phraseology is the branch that studies phrases. The words in phraseologisms (phrases) are units of figurative meaning that combine in meaning, create a meaningful unity and are often equal in content to one lexeme. Phraseologisms are part of a sentence. The words that make up the phraseology are considered only internal parts of the phraseology. It interacts with other parts of the sentence as a whole. The article analyzes issues such as different approaches to phrases, their synonyms, antonyms, the phenomenon of polysemy, and their division into word groups, and examples from works of art are given. Modern, obsolete, dialectal, artistic, scientific, colloquial expressions are also distinguished according to the period and style of use.

Keywords: Phraseology, phrase, fixed phrase, fixed conjunction, proverb, wise word, overall meaning, figurative meaning.

Аннотация: В данной статье анализируются фразеологизмы со специфической семантикой и структурой в языкознании. Фразеология — раздел, изучающий словосочетания. Слова во фразеологизмах (словосочетаниях) — это единицы переносного значения, которые объединяются по смыслу, создают смысловое единство и часто равны по содержанию одной лексеме. Фразеологизмы являются частью предложения. Слова, входящие в состав фразеологизма, считаются лишь внутренними частями фразеологизма. Он взаимодействует с другими частями предложения как единое целое. В статье анализируются такие вопросы, как различные подходы к словосочетаниям, их синонимы, антонимы, явление полисемии, их деление на группы слов, приводятся примеры из художественных произведений. Также выделяются современные, устаревшие, диалектные, художественные, научные, разговорные выражения по периоду и стилю употребления.

Ключевые слова: Фразеология, словосочетание, устойчивое словосочетание, устойчивый союз, пословица, поговорка, мудрое слово, общее значение, переносное значение.

Uzbek folk proverbs and proverbs are a rich socio-cultural, spiritual and educational phenomenon in terms of semantics and linguistic culture. There are in the Uzbek language such linguistic and cultural concepts as "patience", "respect for parents", "childhood", "honor", "kindness", "generosity", "hospitality".

Paremas are ancient wisdoms that have been tested in the experience of people's life and proved themselves over the centuries, and the national-cultural wealth of the Uzbek people in this regard is an incomparable treasure. Not all nations of the world have folk wisdom. In some countries, it has fallen out of use due to neglect. It cannot be said that all Uzbek proverbs and phraseology are fully collected. Many paremas are out of use. If the existing ones are not continuously published and activated for wide consumption, we may lose a certain part of our priceless treasure in the near future. [Kholmanova, 2022: 5]

Language units consisting of a stable relationship of two or more words, ready to be brought into the speech process, existing as a possibility in the memory of the speakers are called stable compounds. The most characteristic features of stable compounds are:

1. Presence in language before the speech process: ready input into speech.
2. Totality of meaning.
3. Stability of structure and composition.

Appropriate use of stable conjunctions ensures the beauty of speech, so they are the beauty of our speech. Although stable compounds are common in terms of the stability of their constituent parts, they differ from the point of view of the totality of meaning. Accordingly, stable compounds are divided into the following groups:

1. Figurative expressions.
2. Phrases (phraseologisms).
3. Proverbs and proverbs.
4. Wise words (aphorisms).

Phrases are unique figurative expressions of conclusions drawn by the people based on observation of events in life, evaluation of favorable and unfavorable actions in society, summarization of life experiences. In general, the volume of phraseological meaning is distinguished by its breadth and complexity compared to the volume of lexical meaning. Components that are not found in the meaning of the word are present in the meaning of many phraseologisms. Writers, as a rule, are not satisfied with the selective use of phrases in accordance with the purpose of the image, but change and rework them according to the character, mental state, and lifestyle of the characters. In this way, folk expressions are polished and saturated with new subtleties of meaning. There are various ways of processing folk expressions, ways of giving them new meanings. This can include methods such as opening a new interpretation of the meaning of a common language phrase, changing the lexical structure of the phrase and expanding its semantic and stylistic functions, adding new figurative and figurative meanings to the phrase.

Fixed connection - words that have the same lexical meaning, united for a semantic whole that is considered as one word. Words in stable conjunctions cannot be replaced, they are stabilized and rounded based on the law of expressing one meaning [Mahkamov, 2013: 118].

A linguistic unit consisting of the combination of more than one independent lexeme and having a figurative and spiritual nature is called a phraseology: the hair on the top stood on end, the vinegar does not hold water, the stomach is cracked, the cap is half, two sentences in one, to raise to the head and another.

Prof. B.Mengliyev emphasizes that phrases are also referred to as phraseologisms, phraseological unit, phraseological combination, and describes them as follows: "Phraseologisms are similar to combined lexemes, phrases and sentences according to their constituents. However, they live ready and stable in the mind of the language community, like more compound lexemes. In other words, phraseologisms have the nature of generality characteristic of all linguistic units in the language and appear as particularities in speech" [Mengliyev, 2012: 25]

Phraseologisms, like lexical units, are also classified in terms of frequency of use. According to this, universal phraseology (to fail, nevertheless, to come true) and phraseology with a limited application are distinguished. Phraseologisms with limited application have certain differences with the period of use (obsolete and modern) and scope (dialectal, scientific, artistic speech).

Obsolete phraseology: to call alif a club, to run away from aliflaila in one fell swoop, to make dustin aliflom, to shake one's head, to shake one's sleeve,

Dialectal phraseology: to be distracted (to worry), to be disheartened (to be crushed), to perish (to be preoccupied), to be thirsty (to be longing) Scientific phraseology: to look, to the eye to be thrown, to be closely connected, to fit in, a circle of possibilities, a comfortable environment.

Artistic phraseology: to be full of patience, to run away from the breath of the mouth, to be full of joy from the mouth, to look for a shriek from the heart, not to burn the sikh, nor the kebab, to light a nut on the head.

Phraseologisms of colloquial speech: to pick barley raw, to hit the ground, the heart is blood, the eyes are narrow, the eyes are open, the head is open.

Most of the phraseologisms are characteristic of artistic and colloquial speech.

Any unit in the language shows national-cultural peculiarities to one degree or another. But there are such language units that are a unique embodiment of national-cultural concepts. In this

sense, phraseology is a vivid expression of national mentality, culture, lifestyle and long-term observations. As the language is formed and develops in a certain society, in this process it reflects all the national characteristics of the nation that is its creator. As a result, languages differ from each other not only in terms of structure, but also nationally and socially. These differences are evident in language units. A representative of a certain language is also a representative of a certain culture. Any culture has its expression in language. Idioms are especially important in expressing the national-cultural identity of the language. Because expressions were created under the influence of various external factors, such as the way of life, culture, and religious beliefs of the language owners.

Therefore, national-cultural realities can also be found in folk sayings: *nonini tuya qilmoq*, *nonini yarimta qilmoq* expressions in relation to misappropriation of someone's rights, *to'nini teskari kiymoq* and the phrase is used to be obstinate while being turned off.

There are several expressions related to the unit of reality:

Do'ppisi tor kelmoq - to get into a difficult situation;

Do'ppisi yarimta - there was joy without sadness;

Do'ppisini osmonga otmoq - to be very happy.

"*Do'ppini olib kel*" *desa, sallasi (kallasi) bilan olib kelmoq* - to do something rougher than it should be. Due to the fact that the turban component in this phrase is not currently used, it was replaced by the lexeme *do'ppi*, as a result of which a new phrase was formed in the form of *do'ppi ol*. If the word *doppi* is used instead of the word *salla*, the word *kalla* is changed to the word *head*. The word *kalla* rhymes with *turban*, providing harmony in tone. As a result of the change of *salla* to *do'ppi*, the initial word monand was chosen for the lip consonants in *do'ppi*: By adding to *do'ppi*, not only the head, but also the limbs are minced (Tahir Malik). In this sentence, it can be seen that the phrase "take the *salla*" has been creatively changed.

Phrases related to the augmented reality unit are as follows:

Og'zidan oshini oldirmoq - to give the right.

Yoshini yashab, oshini oshab - old man or woman.

Social-historical events, moral and spiritual-cultural norms, religious ideas, national traditions and customs, cultural stereotypes and archetypes are reflected in the stable combinations of each language. , they are passed down from generation to generation.

We can find many Uzbek folk expressions in the works of Abdulla Qadiri. For example, in the story "Demon's Party" the following phrases are skillfully used:

"*I barely got rid of the stalk when it was dark and dark*" [Qadiri, 2004: 5].

The phrase "qosh qoraymoq" is used in the sense that it is getting late, it is getting dark in the evening.

I stood there for some time in the morning [Qadiri, 2004: 6].

The expression "tong qotmoq" is used in the sense of becoming unconscious from surprise or fear. The author is here *dong* the word *morning* in the form of

They did not leave me to my own devices, they dragged me and dragged me into the net. Usually, the blessing was recited after entering the assembly and killing [Qadiri, 2004: 12].

In our people, "fotiha o'qimoq" means "duo qilmoq".

Fatiha means "opening, beginning, introduction" in Arabic. 1. As a religious term, it refers to the beginning of the first, initial sura of the Qur'an. 2 A prayer that is said to wish goodness to someone, expresses praise. The word *Fatiha* has undergone a change of meaning in connection with the Qur'anic sura. This surah also expresses the prayers and intentions said without reading. From the word *fatiha*, the compound verbs to read *fatiha*, to bless, to give *fatiha*, to receive *fatiha*, and to apply *fatiha* to one's face are formed:

Fotiha o'qimoq: 1) read one of the short surahs of the Qur'an and pray for God's mercy and goodness for the soul of the deceased; 2) to wish and pray for good things, goodness, peace, health for oneself or others by opening hands and pulling them to the faces.

Fotiha qilmoq colloquial: read the blessing.

Fotiha bermoq - permission, permission given to a person who is going to do something, wishing good luck and happiness.

Fotiha olmoq - acceptance of the blessing of others by the host of the wedding ceremony.

Fotiha so'ramoq - to ask the permission of the meeting people or an older person to get up before others.

Yuziga fotiha tortmoq - after reciting the blessing, gently stroke the face from top to bottom with the palms of both hands.

Oq fotiha The phrase means "permission, permission given with pleasure to start doing something".

Fatiha has become a value as an ethnonym: a pre-marriage ceremony, after a girl and a boy agree to marry, to inform relatives and community members about it; also represents engagement. A small wedding held on the occasion of the engagement of a girl and a boy; an engagement wedding is called a blessing wedding.

A contradictory concept has been formed in the semantics of the word fatiha. This phenomenon is called antisemia. A ceremony is held in the house of the deceased in the first three days of death and on the days of Eid for the memory of the deceased; also means condolence. The words fatihali, fatihakhon, fatihakhanlik, fatihachi are also derived from the word Fatiha. All these units are formed based on the semantics of Surah Fatiha.

Phrases are a spiritual treasure that enriches the language and its content. They are a stable combination that is considered one of the main objects of linguistics, and as a linguistic-cultural object, a language unit that includes the culture of the nation is considered. Broad analysis of expressions from the linguocultural point of view is defined as a problem waiting for its solution in linguistics. Studying them makes it possible to get various new conclusions.

References:

1. Abdullah Kadyri. Bygone Days. – Tashkent, 2004.
2. Abdullah Kadyri. Feast of Demons: Stories. – Tashkent. 2004.
3. Kasimova Z. Analysis of uzbek language household realities in “Bygone Days” / Views of uzbek thinkers on the theory of language, 2021. P.446.
4. Kasimova Z. On the state of some religious realities in the translation process in the novel “Bygone Days” / Current research journal of philological sciences, 2021. P.12-16.
5. Kholmanova Z. and others. Study of concepts related to linguoculturalism. –Tashkent, 2018.
6. Kholmanova Z. and others. Study of Paremiyas in Uzbek language. – Tashkent: 2022.
7. Makhamov N., Ermatov, I. Annotated dictionary of linguistic terms. –Tashkent, 2013.
8. Mengliyev B. Current Uzbek literary language / study guide. – Tashkent, 2012.
9. Rahmatullayev Sh. An explanatory phraseological dictionary of the Uzbek language. – Tashkent, 1978.

THE FORMATION AND DEVELOPMENT OF LINGVODIDACTICS

Fattakhova D.D.

PhD, National University of Uzbekistan.

Mirvaliyeva D.A.

Teacher of the National University of Uzbekistan

Summary: Computer linguistics, also known as computational linguistics, is an interdisciplinary field that merges linguistics and computer science to enable machines to understand, interpret, and generate human language. This area of study has gained significant prominence due to the increasing reliance on natural language processing (NLP) technologies across various industries, from healthcare to finance, and the rise of AI-driven applications like chatbots and language translation systems. Understanding the core issues within computer linguistics is essential for appreciating its current applications and future potential. The field has evolved through distinct phases, beginning with rule-based symbolic approaches in the mid-20th century, transitioning to statistical models in the late 1980s, and culminating in the adoption of deep learning techniques in recent years. Each of these eras has contributed to significant advancements in NLP capabilities, allowing for more nuanced language understanding and processing. However, key challenges persist, including ambiguity in language, limited contextual understanding, and computational complexity, which continue to hinder the development of robust NLP systems. Controversies in the field often center around ethical considerations, such as algorithmic bias and data privacy issues, particularly as NLP technologies become more integrated into everyday life. Researchers and practitioners are increasingly focused on addressing these concerns while ensuring that NLP systems remain effective and equitable. As the field advances, integrating symbolic reasoning with statistical and neural approaches has emerged as a promising direction, fostering the development of interpretable and accountable models capable of tackling complex linguistic challenges. The significance of computer linguistics lies not only in its theoretical contributions but also in its practical applications that enhance communication, improve business efficiency, and facilitate data analysis. As advancements continue, the interplay of technological innovation, ethical considerations, and linguistic research will shape the future landscape of this dynamic field.

Historical Background

The field of computer linguistics, or computational linguistics, has evolved significantly over the decades, marked by distinct eras that reflect the changing methodologies and technologies employed in the analysis and synthesis of human language. The early approaches, which dominated from the 1950s to the late 1980s, were primarily symbolic and rule-based, focusing on the use of defined linguistic rules to parse and understand language.¹³¹⁴

This symbolic era laid the groundwork for foundational theories in linguistics, with key contributions from figures like Alan Turing and Noam Chomsky. The transition to the statistical era in the late 1980s to early 2000s marked a paradigm shift in the field. Researchers began to embrace statistical models that could learn from data, leading to more flexible and powerful natural language processing (NLP) techniques.¹⁵¹⁶

This era saw significant advancements as statistical methods allowed for the processing of large corpora of text data, enabling more accurate language models that were capable of handling the complexities of human language. Following this, the neural network era, which spanned from

¹³ <https://www.cdomagazine.tech/branded-content/from-turing-to-chatgpt-what-the-history-of-nlp-teaches-us-about-the-business-applications-of-ai>

¹⁴ <https://www.linkedin.com/pulse/navigating-evolution-natural-language-processing-nlp-from-prasanth-m-35vrc>

¹⁵ <https://www.linkedin.com/pulse/transition-statistical-methods-nlp-sujit-dhanuka-biopf>

¹⁶ <https://medium.com/nlplanet/a-brief-timeline-of-nlp-bc45b640f07d>

the late 1990s to the late 2010s, introduced deep learning techniques into the mix. This period was characterized by the development of more sophisticated algorithms that could automatically learn linguistic patterns from vast amounts of unstructured data, marking another critical milestone in the evolution of NLP.¹⁷¹⁸

With the advent of large language models (LLMs), the field has further expanded, offering a plethora of applications ranging from chatbots to automated translation systems. Throughout these developments, challenges have persisted. Issues such as limited contextual understanding, ambiguity, and computational complexity continue to pose hurdles for NLP practitioners.¹⁹²⁰

Nonetheless, these challenges also present opportunities for innovation, pushing researchers and practitioners to refine and enhance the capabilities of computational linguistics. Understanding the historical progression of these approaches is vital for appreciating the current state and future potential of the field.²¹

Core Areas of Study

Natural Language Processing (NLP) encompasses a variety of core areas that address the complexities of human language through computational methods. These areas include, but are not limited to, sentiment analysis, question answering, information extraction, and the redaction of sensitive data.

Sentiment Analysis

Sentiment analysis is a significant aspect of NLP that identifies the emotional tone behind a body of text. This technique is widely employed by businesses to gain insights into customer opinions and emotions regarding their products or brands. By analyzing data from social media and customer reviews, companies can detect trends in customer sentiment, which can inform product development, marketing strategies, and customer service improvements.²²

Question Answering

Question answering (QA) is another critical task within NLP that involves receiving a question as input and returning a relevant answer. The simplest form of QA utilizes document retrieval techniques, which find and return matching entries from a knowledge base.²³²⁴ This area of study is vital for creating systems that can interact more naturally with users, providing immediate and relevant information.

Information Extraction

Information extraction focuses on the process of converting unstructured or semi-structured data into a structured format that is easier to analyze. This involves identifying and categorizing key information from textual sources, which can include everything from extracting specific entities to summarizing entire documents. Techniques such as Named Entity Recognition (NER) play a crucial role in this area, allowing for the identification of sensitive data points such as names, addresses, and other personally identifiable information (PII).

Data Redaction

A specialized application of NLP is the redaction of sensitive information, particularly in industries such as finance, insurance, and healthcare. By employing techniques like NER, organizations can effectively identify and de-identify sensitive entries, ensuring compliance with data protection regulations.²⁵

Other Core Areas

In addition to these primary focus areas, NLP also grapples with challenges such as syntax and grammar complexity, where the intricate structure of sentences can hinder the development of

¹⁷ https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4807782_code4425638.pdf?abstractid=4807782&mirid=1

¹⁸ <https://fastdatascience.com/natural-language-processing/guide-natural-language-processing-nlp/>

¹⁹ <https://www.oxfordbibliographies.com/abstract/document/obo-9780199772810/obo-9780199772810-0046.xml>

²⁰ <https://arxiv.org/html/2401.11972v2>

²¹ <https://jair.org/index.php/jair/article/download/15195/26998/37495>

²² <https://arxiv.org/html/2502.11856v1>

²³ <https://link.springer.com/article/10.1007/s11042-022-13428-4>

²⁴ <https://www.quora.com/What-are-the-key-challenges-in-natural-language-processing-and-understanding>

²⁵ <https://spotintelligence.com/2024/01/25/computational-linguistics/>

accurate models. Issues like language differences and training data adequacy are critical in the NLP landscape, as they influence the effectiveness and efficiency of computational linguistic methods²⁶. Moreover, the semantic analysis process must contend with ambiguity and context-dependence, which are inherent in human communication^{27,28}. These core areas and challenges reflect the ongoing evolution of NLP as it strives to bridge the gap between human language and machine understanding.

Major Challenges

Natural Language Processing (NLP) faces several significant challenges that complicate its development and application. One of the most pressing issues is ambiguity in language, where words and phrases can possess multiple meanings depending on their context. This ambiguity can lead to difficulties in accurately interpreting the intended meaning, which is a critical requirement for effective NLP systems^{29,30}.

Data Limitations

Another substantial challenge is data limitations. Working with limited or incomplete datasets can hinder the performance of NLP models, resulting in inaccuracies and inefficiencies in text analysis. The reliance on large, comprehensive datasets is essential for training robust models; however, the availability of such data is not always guaranteed³¹.

Contextual Understanding

Limited contextual understanding is a prevalent issue within NLP systems. Models often struggle to retain and interpret the meanings of words or phrases based on their surrounding context, which can result in misinterpretations of text. Enhancing the ability of models to grasp contextual information is crucial for improving their accuracy and effectiveness in processing language^{32,33}.

Computational Complexity

The computational complexity of NLP tasks presents an additional barrier. Processing extensive amounts of text data requires significant computational resources and time, making real-time analysis challenging. Researchers continue to explore ways to improve the efficiency and speed of NLP algorithms to address this issue^{34,35}.

Disambiguation Techniques

To tackle the ambiguity and complexity challenges, NLP employs various disambiguation techniques. These may include using context clues and statistical models to resolve ambiguities in syntax and semantics. However, developing effective disambiguation strategies remains a significant hurdle for most NLP tasks, as it requires a nuanced understanding of context and meaning^{36,37}.

Applications

Natural Language Processing (NLP) has a wide range of applications across various industries, leveraging its ability to understand and manipulate human language. These applications can significantly enhance business efficiency, customer experience, and data analysis.

Social Media Monitoring

NLP techniques are employed in social media monitoring tools to extract mentions of brands, products, or services from user-generated content. Once these mentions are identified, they

²⁶ <https://i2group.com/articles/the-10-biggest-issues-facing-natural-language-processing>

²⁷ <https://fiveable.me/introduction-linguistics/unit-13/fundamentals-computational-linguistics/study-guide/nBQMft0bL6VpDNLm>

²⁸ <https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1353>

²⁹ <https://www.atltranslate.com/ai/blog/natural-language-processing-nlp-problems-solutions>

³⁰ https://precog.iit.ac.in/pubs/AnmolGoel_Thesis_IIT.pdf

³¹ <https://ns4nlp-coling.github.io/>

³² <https://arxiv.org/html/2312.11651v1>

³³ <https://www.sciencedirect.com/science/article/pii/S2589004224021035>

³⁴ https://people.orie.cornell.edu/davidr/or474/nn_sas.pdf

³⁵ <https://www.xcubelabs.com/blog/hybrid-models-combining-symbolic-ai-with-generative-neural-networks/>

³⁶ <https://direct.mit.edu/coli/article/47/4/707/107177/Natural-Language-Processing-and-Computational>

³⁷ https://www.researchgate.net/publication/357152309_Patterns_of_Lexical_Ambiguity_in_Contextualised_Language_Models

can be analyzed for sentiment, engagement, and other important metrics, which can then inform marketing strategies or evaluate their effectiveness in real-time³⁸.

Business Analytics

The integration of NLP in business analytics provides organizations with the capability to make sense of vast amounts of unstructured data. By transforming this data into actionable insights, businesses can conduct market research, analyze feedback, and implement strategic adjustments based on comprehensive data visualization.

Healthcare

In the healthcare sector, NLP offers numerous benefits, including the extraction of insights from medical records and the analysis of unstructured data. It can enhance clinical decision support systems, optimize chatbot responses for patient care, and monitor adverse drug reactions, contributing to more effective and personalized healthcare solutions.

Fintech

NLP is transforming the financial technology (fintech) industry by enabling seamless document processing, improving risk management, and enhancing fraud detection. It also plays a crucial role in assessing individuals' creditworthiness and personalizing financial products to better meet customer needs.

Media and Advertising

Media and advertising professionals utilize NLP for content personalization and the delivery of tailored messages. This technology aids in the precision analysis and targeting of user personas, enhancing the effectiveness of advertising campaigns.

AI Assistants and Chatbots

AI-powered applications such as Siri, Alexa, and various chatbots utilize NLP techniques to facilitate natural language conversations with users. These systems can respond intelligently to user queries, providing assistance and information in real-time.

Language Translation

NLP models enable accurate language translation by leveraging vast amounts of bilingual data. These models consider contextual nuances and grammatical rules, making them essential for global communication and understanding³⁹.

Email Filtering and Spam Detection

NLP is widely used in email filtering systems, which categorize incoming messages to help users manage their inboxes effectively. Advanced spam detection technologies utilize NLP's text classification capabilities to identify and filter out unwanted messages, enhancing user experience⁴⁰.

Integration of Approaches

The integration of various computational linguistics approaches has become a critical focus in the development of natural language processing (NLP) systems. There are numerous methodologies, each with its contributions and limitations, that need to be balanced effectively to achieve optimal results in language understanding and generation.

Symbolic and Statistical Approaches

One of the foundational strategies in NLP is the combination of symbolic reasoning and statistical learning. The symbolic approach relies on explicit rules and structured representations of knowledge, while statistical methods utilize data-driven techniques to infer patterns from large corpora of language use. This duality allows for a more robust understanding of language, as both approaches complement each other's strengths and weaknesses⁴¹. For instance, neural-symbolic methods are emerging as a way to harness the adaptive learning capabilities of neural networks while maintaining the clarity and logical rigor associated with symbolic AI.

³⁸ <https://levity.ai/blog/11-nlp-real-life-examples>

³⁹ <https://www.netguru.com/blog/neurosymbolic-ai>

⁴⁰ <https://lumenalta.com/insights/13-natural-language-processing-algorithms>

⁴¹ <https://direct.mit.edu/books/edited-volume/4226/The-Balancing-ActCombining-Symbolic-and>

Development of Neuro-Symbolic Computing

Neuro-symbolic computing represents a significant advancement in the integration of symbolic AI and neural networks. This method aims to merge machine learning with reasoning in a coherent manner, thus providing systems that are both explainable and accountable⁴². By leveraging the strengths of both paradigms, researchers are able to create models that not only learn from data but also offer insights into their decision-making processes, enhancing user trust and system transparency⁴³⁴⁴.

Historical Context and Practical Applications

The integration of approaches in computational linguistics has historical roots that date back to the 1960s. One notable early application was Joseph Weizenbaum's ELIZA program, which combined linguistic rules with the illusion of understanding to engage users in conversation⁴⁵⁴⁶. The methodologies developed during this era laid the groundwork for contemporary practices that remain relevant in various business contexts today. As we analyze the evolution of NLP, it is essential to recognize the enduring applications of these integrated approaches in developing systems that can effectively understand and interact with human language.

Current Trends and Future Directions

Natural Language Processing (NLP) is experiencing rapid advancements, driven by innovative research and the development of sophisticated models. One prominent trend is the emergence of retrieval-augmented generation (RAG) models, which combine retrieval mechanisms with generative capabilities. Hugging Face has developed such a model, while other significant models like REAL by Google and DPR by Facebook also contribute to this space⁴⁷⁴⁸.

Advances in Model Efficiency

Research is increasingly focused on enhancing the efficiency of current NLP models. Various architectural variants of transformers are being introduced, designed to process sentences faster and more effectively⁴⁹. The continual refinement of these architectures aims to improve the performance and accessibility of NLP technologies across different applications.

Multi-Modal Learning

Another exciting development is the rise of multi-modal models that integrate text and image processing. This approach is being explored within the Hugging Face library, aiming to introduce common sense reasoning capabilities by combining diverse data types. This integration allows for richer contextual understanding and has potential applications in fields ranging from automated content creation to enhanced search functionalities.

Neuro-Symbolic Approaches

Neuro-symbolic learning represents a critical area of focus, merging the adaptive learning properties of neural networks with the clarity of symbolic AI. This combination seeks to leverage the strengths of both paradigms to address complex NLP challenges, such as knowledge representation and reasoning⁵⁰. As researchers explore these hybrid methods, there is growing interest in their applications to real-world NLP problems, indicating a significant shift towards more interpretable and robust models.

Ethical Considerations and Challenges

As NLP technology continues to evolve, ethical considerations remain paramount. Issues such as algorithmic bias, privacy concerns, and the potential for misuse of data are increasingly scrutinized. Organizations are encouraged to prioritize transparency and accountability in their

⁴² <https://digitalcommons.lindenwood.edu/cgi/viewcontent.cgi?article=1610&context=faculty-research-papers>

⁴³ <https://www.quora.com/Why-is-it-difficult-to-integrate-neural-networks-with-symbolic-reasoning>

⁴⁴ https://journalwjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-0287.pdf

⁴⁵ <https://aclanthology.org/2022.nlpower-1.6/>

⁴⁶ <https://www.engati.com/blog/computational-linguistics>

⁴⁷ https://www.reddit.com/r/linguistics/comments/fyrb65/computers_still_struggle_to_understand_text/

⁴⁸ <https://pmc.ncbi.nlm.nih.gov/articles/PMC7905448/>

⁴⁹ <https://repo.journalnx.com/index.php/nx/article/download/4456/4240/8591>

⁵⁰ https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00670/121540/Scope-Ambiguities-in-Large-Language-Models

NLP initiatives to ensure that these technologies promote inclusive and equitable outcomes⁵¹. Addressing these ethical challenges is vital for the sustainable development of NLP applications.

Future Prospects

The future of NLP is poised to explore cross-lingual transfer capabilities, where word embeddings of different languages can be projected into a shared space, enabling zero-shot learning and unsupervised machine translation⁵². This development has the potential to enhance accessibility for low-resource languages and broaden the applicability of NLP technologies globally.

One of the urgent issues is to increase the authority and status of the Uzbek language as the state language, to strengthen the status of the state language and to raise the level of developed languages of the world. In the Decrees and Resolutions of the President of the Republic of Uzbekistan, resolutions of the Cabinet of Ministers of the Republic of Uzbekistan, special attention is paid to the issues of teaching the Uzbek language in measures aimed at the development of the Uzbek language and increasing its authority as the state language. The Decree of the President of the Republic of Uzbekistan dated October 21, 2019 No. UP-5850 "On measures to radically increase the authority and status of the Uzbek language as the state language" states that one of the main issues is the further improvement of the system of teaching the state language in educational organizations, increasing its authority as the language of science⁵³.

The main issue on the agenda is the replacement of the Uzbek language among the world languages, bringing it to the level of protected and developed languages.

Improving the teaching of the Uzbek language, improving the quality of language teaching, mastering the Uzbek language as a native language, learning as a foreign language is related to computer linguistics.

Linguodidactics is an innovative form of language education, which primarily describes the language for the purpose of teaching. In teaching language capabilities, phonetic, lexical, grammatical features, it differs from traditional language education in that it takes into account the student's age, abilities, level of thinking.

Linguistics has been formed and developed in the field of foreign linguistics. The use of achievements of linguistics in Russian and English linguistics and the study of traditional pedagogical experience resulted in the formation of Uzbek linguistics.

There are two different directions of linguistic didactics: the development of the language as a mother tongue and the study of the language as a foreign language. The literature on linguistics mainly focuses on the study of the language as a foreign language.

Computer linguistics, which emerged as one direction of linguistics, and is developing in a separate field in the next, deals with the problems of organization of language learning with the help of computer technologies.

In world linguistics, the issue of language teaching is a matter at the level of public policy. The use of computer technology in language education, characterization for the purpose of language instruction plays an important role.

Emphasis on the teaching of international languages and the languages of developed countries of the world serves to enhance the position of these languages. The appeal of the great thinker Alisher Navoi "Attention to the language" has retained its value today. The more a nation focuses on its language, its status and its development, the more powerful a nation will be. The experience of the world is proof of this.

One of the first factors in the elevation of the position of the language is the issue of language teaching. Today, the improved methods of teaching Russian, English, German, French, Korean, Arabic and Chinese languages are gradually penetrating into the framework of other

⁵¹ <https://medium.com/@mubarakdaha/enhancing-translation-and-sentiment-analysis-with-machine-learning-a1267d8d1c9c>

⁵² <https://www.studocu.com/en-us/document/university-of-maryland-baltimore-county/historical-linguistics/lecture-12-computational-linguistics/55606916>

⁵³ Decree of the President of the Republic of Uzbekistan UP-5850 dated October 21, 2019 "On measures to radically increase the authority and status of the Uzbek language as the state language".

languages. In this situation it is important to ensure the place of the Uzbek language among the world languages, to preserve the language as a native language, the language of the people, the language of the nation, to expand the scope of its application.

The combination of foreign experience and national pedagogical achievements in the formation and development of the sphere of linguistic didactics as a modern form of Uzbek language education has practical value.

Computer Assisted Language Learning (CALL), English or Computer-Aided Language Learning (CAI)/Computer-Aided Language Teaching (CALI), American, briefly described in Levy's scholarly work. As "Seeking and Exploring the Application of Computers in Language Teaching and Learning". CALL encompasses a wide range of information and communication technology applications and approaches to the teaching and learning of foreign languages, from the "traditional" exercises and practical programs that characterized CALL in the 1960s and 1970s, to the most recent versions of CALL, for example, as used in virtual learning environments and web-based distance education. It can also be found using corpora and concordances, interactive whiteboards, computer-mediated communication (CMC),[4] virtual worlds, and mobile device-based language learning (MALL). The term CALI (Computer Aided Language Router) was used before CALL, its origins being reflected as a subset of the generic term CAI (Computer Aided Language Router). However, CALI has not been adopted by language teachers because it implies a teacher-centered approach (teaching), whereas language teachers prefer to focus on learning rather than teaching. CALL began to replace CALI in the early 1980s (Davies & Higgins 1982: p. 3), and it is now included in a growing list of professional associations around the world. In the early 1990s, an alternative term, technology-enhanced language learning (TELL), also appeared: for example. TELL Consortium Project, University of Hull.

The current CALL philosophy places a strong emphasis on student-centered materials that allow students to work independently. Such materials may or may not be structured, but they typically include two key features: interactive learning and individual learning. CALL uses tools to help teachers facilitate the process of language learning. They can be used to reinforce what has already been learned in the classroom or to help students who need extra help.

References:

1. Abduhamidova L. Automatic editing direction of computer linguistics. BMI. –Tashkent, 2016. 67 b.
2. Абдурахмонова Н. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти. Фил. фан... (PhD) diss. – Тошкент, 2018.
3. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
4. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus. uz. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.
5. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE.
6. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
7. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language

- Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
8. Аврорин В.А. Проблемы изучения функциональной стороны языка (к вопросу о предмете социолингвистики). –Л., 1975. –С. 177.
 9. Адилова С. Ўзбек тилини ўқитишда компьютер технологияларидан фойдаланиш. – Тошкент, 2006.
 10. Акопянц А.М. Прагматика и лингводидактика. –М., 2008.
 11. Быкова С.Е. Тезаурусное моделирование юридической терминологии (на материале французского языка в сопоставлении с русским). – Ekaterinburg, 2006.
 12. Suyunov B. Semantics and thesaurus of medical terms. Philol. Fan. d-ri (DSc) ... Diss. autoref. –Tashkent, 2023.
 13. Kholmanova Z. Computational linguistics. Study guide. Tashkent , 2019.
 14. Oxford Dictionary of English (oed.com)
 15. Collins Automatic Annotated Dictionary of English (mycobuild.com)
 16. Automatic version of the dictionary "New Large-Size English-Russian" by D. Apresyan D. and E.M. Mednikova (<http://eng-rus.slovaronline.com>)
 17. Ojegova online dictionary (<http://slovarozhegova.ru>)
 18. Yu and L. Deng, "Automatic Speech Recognition: A Deep Learning Approach" (Publisher: Springer), 2014.
 19. Анисимов А.В. Компьютерная лингвистика: Мифы. Алгоритмы .Язык. -Киев: Наукова-думка, 1991, с.15
 20. www.Udemy.com
 21. www.coursra.com

ОСОБЕННОСТИ ПОСТРОЕНИЯ И ФУНКЦИОНИРОВАНИЯ АВТОРСКИХ КОРПУСОВ

Джуманиязова Интизор

Преподаватель университета Маъмун

E-mail: intizor1828@gmail.com

Аннотация: В статье рассматриваются актуальные вопросы, связанные с развитием авторской лексикографии и корпусной лингвистики, а также их взаимосвязь. Подчеркивается важность создания авторских корпусов для более глубокого анализа языковых особенностей произведений известных писателей и поэтов. Рассматриваются примеры авторских корпусов Ф.М. Достоевского и А.С. Пушкина. Корпус произведений А.С. Пушкина стал значимым этапом в изучении языка автора. Анализируются проблемы, связанные с грамматической разметкой, частотностью слов и омонимией в авторских корпусах. Также указывается необходимость их дальнейшего совершенствования в создании полноценных лексикографических ресурсов.

Ключевые слова: авторский корпус, авторская лексикография, конкорданс, электронный словарь, корпус Пушкина, Достоевского.

Abstract: The article discusses current issues related to the development of author's lexicography and corpus linguistics, as well as their relationship. The importance of creating author's corpora for a deeper analysis of the linguistic features of the works of famous writers and poets is emphasized. Examples of the author's works by F.M. Dostoevsky and A.S. Pushkin are considered. The corpus of A.S. Pushkin's works became a significant stage in the study of the author's language. The problems related to grammatical markup, word frequency, and homonymy in author's corpora are analyzed. It also indicates the need for their further improvement in the creation of full-fledged lexicographic resources.

Keywords: author's corpus, author's lexicography, concordance, electronic dictionary, corpus of Pushkin, Dostoevsky.

Введение. Авторский корпус тесно связан с авторской лексикографией, поэтому уровень развития лексикографии находится в тесной взаимосвязи с корпусной лингвистикой: развитие одного невозможно без развития другого. К сожалению, эти два направления языкознания медленно завоевывают позиции в мировой лингвистике. Это еще развивающееся направление, в котором, однако проделана определенная работа по развитию авторской лексикографии. Однако сегодня все еще остается много вопросов, ожидающих решения. Для авторских корпусов также наблюдается медленная динамика развития. Некоторые исследователи обозначают это более оптимистично, определяя развитие «стабильным», связывая с расширяющимися возможностями языковых корпусов [Абдурахманова Н.З., 2021].

Существующие авторские корпуса (корпуса европейских языков) имеют довольно простую разметку. Однако хорошо структурированный, глубоко проиндексированный корпус авторских работ может сегодня способствовать развитию лексикологии, лексикографии, истории языка, диалектологии, социолингвистики и других областей, открывая двери новым возможностям [Хидирова Г.Н., Кадилова З.Р., 2023]. Примером тому может служить создание языкового конкорданса известного писателя/поэта с использованием корпуса авторов, а также создание Национального корпуса и базы данных.

Основная часть. Достижения корпусной лингвистики русского языка в этом отношении заслуживают внимания. Корпус произведений великих писателей и поэтов, таких как А.П. Чехов, А.С. Пушкин и Ф.М. Достоевского, корпус В. Шекспира в Британском национальном корпусе и корпус И. Гёте в корпусе Мангеймского

университета; Корпус персидско-таджикской классической и современной литературы является достижением корпусной лингвистики и образцом для его дальнейшего развития.

Создание авторского корпуса произведений Пушкина стал уникальным опытом, поскольку помог в уточнении сложных языковых свойств авторских текстов. Проект «Пушкинский язык: электронная энциклопедия» стал значимым шагом в этом направлении. Известно, что язык Пушкина играет основополагающую роль в современном русском литературном языке, влияя на формирование грамматических норм. В связи с этим творчество Пушкина стало предметом более глубокого и детального анализа, что способствовало лучшему пониманию языковых особенностей.

Существующие электронные издания произведений Пушкина, которые отражают языковую природу, представляют собой лишь поисковую систему, основанную на словоформах, а также незначительное количество аннотаций и иллюстраций, созданных из электронного набора. Потребность в разработке лингвистической базы данных возникла вследствие того, что язык его произведений до сих пор не получил комплексного и всестороннего исследования. Лексика и фразеология Пушкина остаются недостаточно охарактеризованными, также как и не систематизированы грамматико-морфологические, морфемные, словообразовательные и синтаксические особенности его языка. Исследование специфики использования слов и словосочетаний в произведениях Пушкина различных жанров как целостной системы остается незавершённым.

Создание корпуса сочинений А.С. Пушкина предполагает проведение комплексного лингвистического анализа, который включает в себя грамматическую и семантическую классификацию слов и морфем в текстах автора. Произведения Пушкина, обладающие яркими и разнообразными характеристиками, должны быть доступны для изучения специалистам, что позволит им проводить самостоятельные исследования. При этом важно наличие развитой базы данных, что поспособствует формированию научных выводов. Уникальной чертой корпуса является то, что даже само слово, будучи основой, содержит богатую лингвистическую информацию, открывающую новые горизонты в области языкознания и истории языка. База данных будет служить новым этапом в исследовании особенностей языка Пушкина и станет основой для изучения лексики, синтаксиса и структуры его текстов.

А.А. Поликарпов, обсуждая создание "Электронной энциклопедии языка Пушкина"[Поэзия и драматургия А.С. Пушкина. Путеводитель по Пушкину. Вып. 1. - М., 2006], выражает сожаление по поводу отсутствия аналогичных изданий для других авторов. Он отмечает медленный процесс развития авторской лексикографии. Поликарпов считает, что область авторских корпусов остается не до конца реализованной в международной корпусной лингвистике [Научный семинар, 2011]. Анализ существующих авторских корпусов показывает, что они отражают лишь небольшой аспект языковой характеристики конкретного писателя. Электронные корпуса чаще рассматриваются как узкофункциональные и в первую очередь информативные. Таким образом, наблюдается потребность в более глубоком исследовании и развитии авторских лексикографических ресурсов для полноценного понимания литературного языка.

Мировая компьютерная лексикография также сталкивается с нерешенной проблемой. Главное (и, по сути, единственное) средство поиска в глобальных авторских корпусах остается неопианным — это отсутствие четкого обозначения омонимов недифференцированного слова и формы [Хамраева Ш.М. DSc Ф.- 2021].

Корпус текстов Федора Достоевского был разработан с целью создания словаря, который служит основой для лексикографического анализа. В основе системы лежит структура, содержащая различные тексты (произведения, письма), что позволяет охватить полное разнообразие словарного запаса писателя. В состав корпуса входят все его художественные и публицистические произведения, а каждое произведение рассматривается как отдельная единица хранения текста, начиная с романов и заканчивая

отдельными буквами. В настоящее время вся художественная и документальная литература, а также эпистолярные материалы Достоевского доступны на электронных носителях; работа хранится в формате ASCII, с минимальным тегированием для указания страниц.

На основе авторского корпуса Ф. Достоевского были разработаны различные словари, включая частотный словарь, составленный А.Я.Шайкевичем. Идиоматический словарь, отражающий авторские идиомы, также структурирован. В дополнение к этому, корпус представлен на компакт-диске «Достоевский: тексты, исследования, материалы».

В словаре-конкордансе Достоевского каждое слово в его текстах (литература, публицистика, авторский «дневник», представленных в 30-томнике его замечательных произведений) сопровождается контекстом использования и частотностью появления, эти параметры соединены с функцией перехода/обращения к основному тексту. В качестве материала взяты все завершённые произведения, а также незавершённые и те, чья авторская принадлежность вызывает сомнения (например, «Ряд статей о русской литературе» - «Вопрос об университетах») [Хамраева Ш.М. Международный журнал-2020 год]. Под редакционными комментариями, ссылками, авторскими заметками и черновиками подразумеваются произведения, не включённые в тома. В отличие от корпуса А.Чехова, в коллекции собранных трудов Достоевского редакторские комментарии были также учтены как составная часть корпуса.

Этот словарь-конкорданс стал основным ресурсом для частотного анализа произведений Достоевского, позволяющего исследовать частоту использования слов. На этапе подготовки текста основным элементом для корпуса служит слово, а именно его форма и лемма, которые помогают определить частоту слов, используемых автором как в оригинальных фразах, так и в цитатах. Словосочетания, состоящие из собственных существительных, группируются в леммы как единые конструкции (например, журнал "Русский вестник" трактуется как название; аналогичный процесс применяется для проведения поиска.). Это позволяет установить, с какой частотой автор обращается к различным изданиям и произведениям. Ещё одним преимуществом конкорданса является возможность анализа контекста слова, что помогает определить его размещение. Пользователь видит искомое слово слева/справа от его формы и может наблюдать контекст, состоящий из десятков слов, что облегчает выделение основного текста. Это оцифровка текста предоставляет возможность сопоставления с изданиями.

Отметим некоторые особенности корпуса. В основном разделе записи представлены заглавными буквами (например: слово "Белинский" в форме "БЕЛИНСКИЙ"), однако при поиске слова форма буквы не имеет значения.

Встречающиеся в первоисточнике выделенные курсивом слова по техническим причинам нейтрализованы.

Лексические омонимы объединены в одной статье, а не представлены отдельно (пользователь может легко определить омонимы в зависимости от контекста).

Составные словосочетания и устойчивые выражения структурированы по смысловому признаку.

Аббревиатуры, сокращённые обозначения, не поясняются, цифры не переводятся в слова.

Часто используемые журналы и наименования газет, записанные в сжатом формате, были подвергнуты лемматизации и аннотации. (Например, Совр. – «Современник», Вест. Европы – «Вестник Европы»).

Слова, содержащие дефис, были нормализованы, например, "без-дар-ного" – «бездарный».

Орфография слов осталась без изменений (сохранена орфографическая форма авторского периода), например, леммы "парке", "Толоза".

Иноязычные слова, которые не являются заимствованиями, не подвергаются лемматизации.

Вывод. На основе анализа принципов составления вышеуказанных авторских корпусов, отметим их уникальные, отличительные и общие аспекты. К общим отнесем: создание различных (алфавитных, частотных, обратных, семантико-грамматических) словарей как производных авторского корпуса, разработка отдельных методик и концепций создания корпуса, подбор источников для корпуса и словаря, разметка корпуса, структура словарной статьи, признак, характеризующий лексему (заглавное слово, краткое пояснение, грамматическая характеристика, частотность) [Зализняк, 1980. 880].

Грамматический признак словоформы и характеристика ситуации употребления слова (контекста) являются основными проблемами при составлении авторского корпуса. Корпус каждого автора имеет свои уникальные характеристики, такие как отбор материала для корпуса (какие тексты автора включены/не включены в корпус), структура словарной статьи (представлен ли статус омонимии и вариантных форм в одной словарной статье или отдельно), интерфейс корпуса и менеджер корпуса.

Список использованной литературы

1. Абдурахманова Н.З. Компьютерная лингвистика: учебное пособие. – Ташкент – 2021.
2. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL), 6(1-2019), 131-7.
3. «Авторская лексикография и история слов: к 50-летию выхода в свет «Словаря языка Пушкина» /Научный семинар. 16-17 декабрь 2011г. Институт русского языка им. В.В. Виноградова.
4. Бобунова М.А. Конкорданс XXI в.: новая старая форма // Вопросы лексикографии. - 2016. №2 (10) - С.41-54.
5. Зализняк А.А. Грамматический словарь русского языка. М.: Русский язык. - 1980. – 880 с.
6. Поэзия и драматургия А.С. Пушкина. Путеводитель по Пушкину. Вып. 1. - М., 2006.
7. Столяров А.И. Словарь-конкорданс и его применение в рамках корпусной лингвистики // Гуманитарные научные исследования. - 2017. № 2. URL: <http://human.snauka.ru/2017/02/21074>
8. “Теория и практика авторской лексикографии” / Материалы семинара. www.ruslang.ru/seminar_aut_lexocorg020413
9. Хамраева Ш.М. Лингвистическое обеспечение морфологического анализатора узбекского языка. DSc Ф.- 2021
10. Хамраева Ш.М. Особенности и общие характеристики авторских корпусов. Международный журнал «Искусство слова», номер 2, выпуск 3. – 2020 год
11. Хидирова Г.Н., Кадирова З.Р. Корпусная лингвистика и художественный перевод. Монография. – Т., Изд. «Иқтисодиёт». – 2023. – 180 стр.
12. Якубовский К.И., Якубовская К.А. Обзор современных лингвистических технологий и систем // Вестник МГУП. - 2015. №2.- С.315-319.
13. www.philol.msu.ru
14. dostcorpus.tilda.ws

O'ZBEK TILI TEZAVURSI SEMANTIK MAYDONLARI VA STRUKTURASI

Xoliyorova Gulruxsor G'ayratovna

Mirzo Ulug'bek nomidagi

O'zbekiston Milliy universiteti tadqiqotchisi

Annotatsiya: Ushbu maqolada tilshunoslikning turli sohalari, jumladan, semantika, ontologiya, va terminologiyaning ahamiyati ko'rib chiqiladi. Shuningdek, O'zbek tili elektron tezavrusi semantik maydonlari va strukturasi yaratishda ontologik va terminologik tamoyillarni qo'llash va ulardan foydalanish bo'yicha takliflar ishlab chiqiladi. Maqolada, milliy lisoniy modelni shakllantirishga qaratilgan konseptual yondoshuvlar hamda xalqaro tajribaga asoslangan usullar haqida fikr yuritiladi.

Kalit so'zlar: Elektron tezavrus, ideografik lug'at, terminologik tezavrus, lisoniy model, semantik maydon, korpus lingvistika.

Kirish

Tezavrusning semantik munosabatlari so'z turkumiga qarab turlicha bo'lishi mumkin. Ushbu termin keng ma'noda qo'llanib, narsa-predmetlarning tur jins munosabatlari ifodalashda umumiy keng ma'noda qo'llanadi. Tilshunoslikda umumiy mohiyatni anglatuvchi boshqa tushunchalar bilan munosabatni anglatuvchi kengroq ma'noga ega tushunchalar yig'indisini bildiradi. Unga tegishli bir turdagi narsa nomlarini atab kelgan tushunchalar, ya'ni torroq ma'noga ega muayyan tur yoki jinsga tegishli so'zlar giponim deb nomlandi. Giponimlar va giperonimlar o'rtasidagi munosabatga tegishli yondashuv giperonimni giponimlardan tashkil topgan deb qarashimizga asos bo'ladi. Giponimlar odatda otlarga nisbatan qo'llanilsa-da, ular nutqning boshqa qismlariga nisbatan ham ishlatilishi mumkin. Ismlar singari, fe'llardagi giperonimlar ham keng turdagi harakatlarni bildiruvchi leksemalar sanaladi. Masalan, ruchka, daftar, qog'oz kabi so'zlar o'quv qurollariga kiradi. Giponimiya va giponimiya tushunchalari yuqoridan quyiga qarab iyerarxik tuzilishda ifodalanar ekan, umumiylikdan xususiylikka qarab ma'no semantikasi torayib boradi.

Leksik munosabatlarning muhim sinfini so'z shakllari va morfologik aloqalari tashkil qiladi. Zotan, WordNetga so'zning muayyan grammatik kategoriyasi bilan birgalikda so'z kiritilganda uning boshlang'ich shaklini topish zarur: Books=>book, geese=> goose; go-went-gone-going kabi. Iyerarxik tarzda so'zlarning tasniflanishi muayyan chegaralarga ega. Masalan, meronim butun qism munosabati faqat ot so'z turkumi doirasida bo'lsa, antonim ham shunday holatda bo'ladi. Shu jihatdan so'zlarning uch jihati inobatga olinadi: atributlari, qismlari (meronim), funksiyalari. Shu jihatdan WordNet turli semantik komponentlarni o'z ichiga olgan yigirma beshta fayldan iborat. WordNetdagi asosiy munosabat bir xil so'z turkumi (ST) orasida hosil bo'lgan. WordNet so'z turkumlaridan iborat ichki guruhga va ayrimlari kesishgan ST (POS) bo'lingan. Ushbu kesishuvchi ST munosabatlari o'zakdosh bo'lgan morfosemantik aloqadagi tizimdan iborat: observe (verb), observent (adjective), observation, observatory (noun). Ot-fe'l juftliklari ot bilan fe'lning semantik munosabatida o'zaro juftlashadi: {sleeper, sleeping car} {sleep} uchun MAKON va {painter} {paint}ning AGENTi, shunda {painting, picture} NATIJA bo'ladi.

WordNetda fe'l so'z turkumiga kiruvchi troponimlar kiritilgan. Unga ko'ra u yoki bu fe'l boshqa muayyan shakliga ko'ra bog'lanadi. Masalan, vaysamoq gapirmoq fa'lining troponimi hisoblanadi. Shuningdek, fe'llar ham antonimik munosabatda bo'lishi mumkin: turmush qurmoq – ajrashmoq.

Shuningdek, lug'atda uyadosh (word similarity) va semantik bog'liq so'zlar (word relatedness) ham lingvistik baza sifatida kiritiladi. Masalan, ananas va o'rik o'zaro sinonim so'zlar emas, lekin ular meva turkumiga kirganligi uchun bitta paradigmaga birlashadi va uyadosh so'zlarni tashkil qiladi. Ularning tarkibida esa ho'l meva va sitrus meva kabi yana ichki turlari

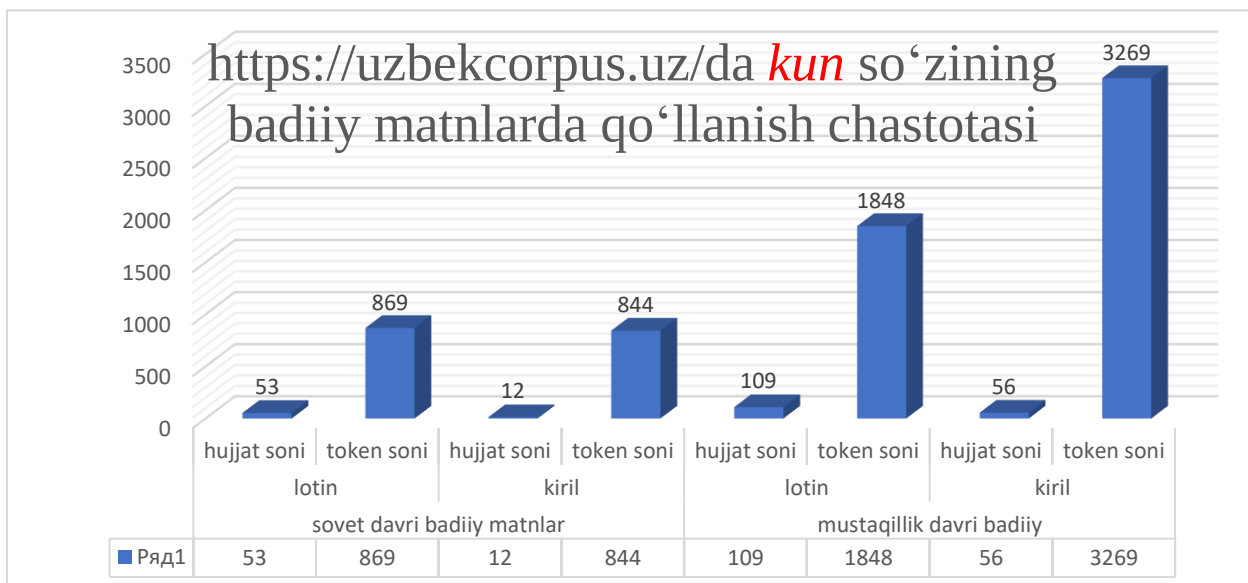
mavjud. O‘zaro semantik bog‘langan so‘zlar esa biri ikkinchisini kelishini talab qiladi hamda assotsiativ xotirada ular hamisha yonma-yon keladi. Masalan, paxta so‘zi bilan ~yig‘moq, ~terimi, ~dek yumshoq kabi so‘zlar bilan valentlik hosil qiladi. Bunda paxta yuqoridagi leksemalarning hech qaysi biri bilan ma‘no o‘xshashligi jihatidan bog‘liq emas, lekin yonma-yon kelganda bir-birini kelishini talab qilib, muayyan semantik guruhga birlashadi.

Endilikda UzNET deb nomlangan o‘zbek tili WordNetining lingvistik bazasi bo‘yicha amaliy ishlar bajarilmoqda. Bunda, asosan, ot, sifat va fe‘l so‘z turkumiga doir so‘zlarning semantik maydonlari beriladi : ishla- leksemasining semantik maydonlari tahlili

1. bosh so‘z: ishla
2. semantik uya: mehnat qilmoq, ter to‘kmoq
3. leksik-semantik guruh: boshqarmoq, haydamoq, qurmoq, yasamoq, yaratmoq
4. semantik to‘da: zavod, korxona, ishchi, dehqon, dala, dastgoh
5. mikromaydon: mehnat, ish
6. makromaydon: ish-harakat
7. sinonim: mehnat qilmoq, ter to‘kmoq
8. antonim: bekor yotmoq, dangasalik qilmoq
9. geperonim: mehnat qilish (faoliyat ko‘rsatish)
10. giponim: ishlamoq
11. xolonim: harakat
12. meronim: qimirlab qo‘ymoq
13. graduonim: timirskalanmoq-ishlamoq-ter to‘kmoq
14. funksionim: yaratish, vujudga keltirish
15. sath: leksik sath
16. so‘z turkumi: fe‘l
17. morfologik o‘zgarishi: o‘timli&o‘timsiz, bo‘lishli&bo‘lishsiz shaklga ega, zamon, mayl, nisbat qo‘shimchalarini oladi (ishladi-ishlatdi; ishladi-ishlamadi; ishladi-ishlayapti-ishlamoqchi-ishla(gin)-ishla(sin), ishlaylik-ishlangiz-ishlasin(lar); ishladi-ishlandi-ishlashdi-ishlattirdi)kabi
18. tuzilishi: sodda, yasama (ish+la)
19. sintaktik aloqa: zavodda ishlamoq, qattiq ishlamoq
20. assotsiatsiya: ishchi, dehqon, xizmatchi, korxona, tez, sekin, dala, zavod
21. ibora: ishlasang-tishlaysan, ishlamasang-tishlamaysan

Korpus tezaurus uchun asosiy lingvistik resurs vazifasini bajaradi. Konkordanslarning n-gram modeli orqali aniqlangan so‘zlarning semantik masofasi orqali leksemalarning bir-biriga yaqinlik darajasini aniqlash mumkin. Masalan, kun so‘zi bilan kelgan konkordanslarni <https://uzbekcorpus.uz/> orqali tahlilga tortsak.

Dastlab korpus orqali konkordanslarning n-gram modeli orqali aniqlash uchun ularni token bo‘yicha statistik aniqladik. Unga ko‘ra adabiy tilga yaqinroq shakllarni tahlil qilish uchun sovet va mustaqillik davrida yaratilgan badiiy asarlarning korpusi orqali ham kiril, ham lotin yozuvidagi manbalardan ushbu so‘zning semantikasini ifodalashda qidiruv interfeysidan foydalanildi:



Endi natijalardan olingan leksik jadvalni tahlil qilamiz:

Bozor kun	W_1	Dam olish kuni maʼnosida
Tunov kun	W_2	Voqealik vaqtiga ishora
Shu kundan boshlab	W_3	Hozirgi zamon belgisi
Oʻsha kun	W_2	Voqealik vaqtiga ishora
Qanday kunlar keldi	W_4	Ijtimoiy holatga ishora
Bugungi kunda	W_3	Hozirgi zamon belgisi (katta vaqt oraligʻi)
Bir kuni	W_5	Voqelikning ehtimollik vaqti
Kun yorishib qolmoq	W_6	Yorugʻlikka ishora
Kun tigʻi	W_7	Quyoshga ishora
Juma kuni	W_8	Haftaning muayyan kuni / islom dini uchun muborak kun
Kun ogʻib	W_7	Quyoshga ishora
Kun qiyomdan oqqan	W_7	Quyoshga ishora
Kun botib	W_7	Quyoshga ishora
Kunlardan bir kun	W_5	Voqelikning ehtimollik vaqti
Begim kuni	W_9	Anʼana kuniga ishora
Kun kechirmoq	W_{10}	Hayot maʼnosida (ibora)
Itning kunini boshiga solmoq	W_{11}	Jazo maʼnosida (holatga ishora)
Kundan-kunga	W_{12}	Vaqtning ortib borish darajasi
Bir kunmas bir kun	W_5	Ehtimollik belgisi
Kun qizigʻi	W_7	Quyoshga ishora
Qirg kun toʻy qilmoq / yeti kun	W_{13}	Udumga ishora
Yomon kun $\Leftrightarrow$ yaxshi kun	W_{14}	Moddiy va maʼnaviy jihatdan nazarda tutilgan holat, kayfiyat
Bir kuniga yaramoq	W_{15}	Asqotmoq (ibora)
Kunini koʻrsatmoq	W_{16}	Jazolamoq (ibora)
Qora kun	W_{17}	Insonning boshiga ogʻir savdolar tushgan holat
Oʻtgan kunlar	W_{18}	Tarix, xotiraga ishora
Kun yuzli	W_{19}	Yorugʻ, issiq kabi sifatga ishora

Kunda / har kuni	W_{20}	Doimiy takrorlanadigan holatlar nisbatan butunning qismi
Og'ir kun	W_{17}	Insonning boshiga og'ir savdolar tushgan holat
Birovga qolgan kun	W_{21}	Birovga tobelanish holati
Qolgan kun	W_{22}	Sinovga berilgan muddat

Tezaurusning semantik maydonlarini shakllantirishda u yoki bu so'z ishtirok etgan kontekstadgi yaxlit butunlik tahlilga tortilishi maqsadga muvofiq. Ana shunda uning konseptlar atrofida ma'no yaqinlik darajasi aniqlanadi. Shu bois o'zbek tili tezaurusini yaratishda ingliz tilidan avtomatik usulda tarjima qilish orqali konsept va ularga tegishli lemmalarni aniqlash qaysidir ma'noda afzallik yaratsa-da, lekin yilning o'ziga xos semantik qamrovi turlicha va buni korpus orqali tahlilga tortish hamda uning atrofida uchragan so'zlarning semantik yaqinlik darajasiga ko'ra maydonlarni hosil qilish maqsadga muvofiq sanaladi.

Adabiyotlar ro'yxati:

1. SAGER, JUAN C. "Terminological Thesaurus. A more appropriate designation or a deprecated synonym?" *Lebende Sprachen*, vol. 27, no. 1, 1982, pp. 6-7. <https://doi.org/10.1515/les.1982.27.1.6>
2. Осокина С. Основания лингвистической теории тезауруса: автореф. док фил.наук, Челябинск, 2015. – С. 25.
3. Ведерникова, Ю. В. Тезаурусный подход к моделированию английской когнитивно-лингвистической терминологии на примере терминополья Blending theory (теория блендинга) // Филологические науки. Вопросы теории и практики. - 2015. - No 12 (54). - С. 47-50.
4. Gorbunova, N. N. Nekotory'e rezul'taty' tezaurusnogo modelirovaniya angloyazy'chnoj terminosistemy' menedzhmenta (Some results of tesaurus modeling of English management terminological system) // *Uchenyye zapiski natsional'nogo obshchestva prikladnoy lingvistiki*. – 2014. - P. 7-17.
5. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
6. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: *uzbekcorpus. uz*. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.
7. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE.
8. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.

O'ZBEK TILIDAGI IJTIMOYIY TARMOQ MATNLARI SENTIMENT TAHLILI

Sa'dullayev Sarvarbek Komil o'g'li

Mirzo Ulug'bek nomidagi

O'zbekiston Milliy universiteti

Jurnalistika va o'zbek filologiyasi fakulteti talabasi

Annotatsiya: Mazkur maqolada o'zbek tilidagi ijtimoiy tarmoq matnlarining sentiment tahlili masalasi yuzasidan atroflicha fikr yuritiladi. Bunda sentiment tahlil texnologiyalaridan foydalanadigan dasturlar tahlil xronikasi misol keltiriladi. Namunalar keltirilib, amaliy tahlil qilinadi. Xususan, Youtube platformasidagi bir videoga bildirilgan fikrlar misolida sentiment tahlilga asoslanib, videoga nisbatan kuzatuvchilarning munosabati o'rganiladi. Sentiment tahlilning jamiyat uchun beradigan foydalari turli sohalar misolida ko'rib chiqiladi. NLP sohasining dolzarbligi mazkur tahlilning bugungi kundagi ahamiyati singari muhim ekanligi e'tirof qilinadi.

Kalit so'zlar: sentiment tahlil, chatGPT, deepseek, binar model, ko'p klassli model, aspektga asoslangan model, dataset, tokenlash, lemmatizatsiya, stemming, stopwords, parser dasturi.

Annotation: This article provides a detailed discussion of the issue of sentiment analysis of social network texts in the Uzbek language. In this article, an analysis chronicle of programs using sentiment analysis technologies is presented as an example. Examples are given and practical analysis is performed. In particular, based on sentiment analysis, the attitude of viewers to the video is studied based on the example of opinions expressed on a video on the YouTube platform. The benefits of sentiment analysis for society are considered using the example of various industries. It is recognized that the relevance of the NLP field is as important as the importance of this analysis today.

Keywords: sentiment analysis, chatGPT, deepseek, binary model, multiclass model, aspect-based model, dataset, tokenization, lemmatization, stemming, stopwords, parser.

Jadal rivojlanib borayotgan texnologiyalar insoniyat hayotiga tabora chuqur kirib bormoqda. Binobarin, ular odamlar orasidagi asosiy muloqot vositasiga aylanib ulgurdi. Oldimizda turgan yana bir masala esa endi texnologiya va inson lingvo aloqasidir. Shuning uchun NLP sohasidagi har qanday yangilik yoki ilgarilash juda dolzarb. Ma'lumki, bugungi kunda sun'iy intellekt dasturlari rivojlanib bormoqda, xususan, chatGPT, deepseek kabi ilovalar o'zining millionlab foydalanuvchilariga ega. Xo'sh ular qanday ishlaydi?

Dastlab dastur NLP texnologiyalaridan foydalanib kiritilgan matnni tushunishga harakat qiladi:

1. Matndagi so'zlar, gap bo'laklari va grammatikani tahlil qiladi.
2. Matnni semantik jihatdan o'rganadi.
3. Agar buyruq yoki savol bo'lsa, uning kontekstini aniqlaydi.

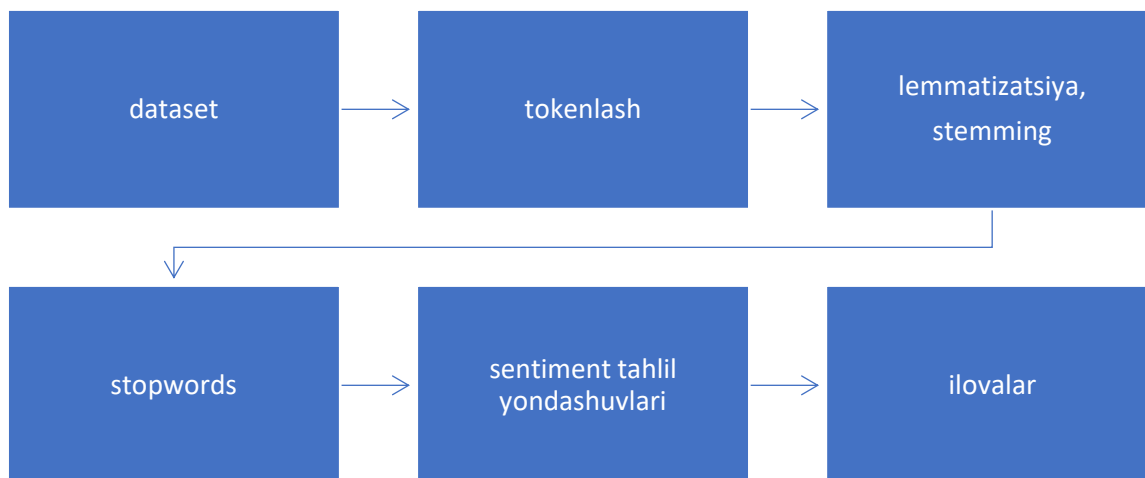
Keyingi bosqichlarda ma'lumotlar qayta ishlanadi va natijani yaratish hamda taqdim etish jarayoni ishga tushadi.

Ko'rinib turibdiki, aqlli dasturlarning foydalanuvchi maqsadiga muvofiq natijalarni taqdim etishida berilgan murojaat yoki buyruqning emotsional bo'yog'ini – sentimental vaziyatini idrok etishi juda ham muhim. Sentiment tahlil bir nechta modellarda amalga oshishi mumkin.

Birinchisi – **binar**. Bunda tahlil qutbli, ijobiy va salbiy jihatlar yuzasidangina amalga oshadi. Misol uchun: Maqola yaxshi ekan. (ijobiy); Maqolada yangi fikrlar yo'q. (salbiy)

Ikkinchisi – **ko'p klassli**. Mazkur model ikkitadan ko'p belgilar bilan ishlay olish imkoniyati bilan ajralib turadi. Misol uchun: Shifokor tajribali. (ijobiy); Shifokor tajribasiz. (salbiy); Shifokor keldi. (neytral)

Uchinchisi – **aspektga asoslangan**. Bu turdagi model kiritilgan matnni ko‘plab xususiyatlar bilan tavsiflay oladi.



O‘zbek tilidagi matnlarni sentiment tahlil qilish bosqichlari dastlab berilgan matnni fundamental NLP qadamlaridan, masalan transliteratsiya, tokenlash, hamda stopwordlardan tozalash kabi qadamlardan o‘tkazib, keyin sentiment tahlil qiluvchi modelga kerakli formatda yetkazib berishdan iborat. Bu bosqichlar yuqoridagi blok-sxemada ko‘rsatilgan⁵⁴.

Dasturning kiritilgan matnni tushunishi jarayonida uni tokenlarga ajratishi muhim sanaladi.

Tokenizatsiya – elektron matndan birliklarni, ya’ni kiritilayotgan belgilar, ramzlar va so‘zlarni ajratib chiqish jarayoni. Mazkur texnologiya grafematik tahrir bosqichida muhim o‘rin tutadi. Lingvistik tahlil dasturlash tillarining manba kodi tarjimalari va kompilyator (o‘zgarlar tadqiqot natijalarining mustaqil yoki qo‘shimcha ishlovlarsiz berilishi)larida, shuningdek, tabiiy tillarning turli parser – dasturlarning LTLarini yaratishda qo‘llaniladi⁵⁵.

Bu o‘rinda parser ham sintaktik analizator vazifasini bajaruvchi maxsus dastur ekanligini eslab o‘tishimiz kerak.

“So‘z asoslarini (stemlarni) topishga qaratilgan jarayon **stemming** hisoblanadi. Stemming stem (asos)larni topishga qaratiladi. Stemmer so‘zni kontekstga bog‘liq bo‘lmagan holda qayta ishlaydi, turli turkumga oid bo‘lgan turli ma’nodagi so‘zlarni farqlay olmaydi. Stemmerlar nisbatan oddiy, ma’lumotlarni tez sur’atda qayta ishlaydi. Lekin ko‘p hollarda hal qiluvchi ahamiyatga ega bo‘lmaydi⁵⁶.

Lemmatizatsiya – morfologik analizning spetsifik masalasi, so‘zning boshqa so‘zshakllaridan kelib chiqadigan dastlabki shakllanish jarayoni. Lemmatizatsiya – so‘zning dastlabki, boshlang‘ich formasini (lug‘atdagi shaklini – lemmasini) tashkillashtirish texnikasi bo‘lib, bu jarayon shu so‘zning boshqa so‘zshakllaridan kelib chiqqan holda amalga oshiriladi⁵⁷.

Yuqorida sanab o‘tganlarimiz foydalanuvchi tomonidan dasturga yuklangan “command” yoki odatiy matnni kompyuterga tushunarli formatga o‘tkazish jarayonining asosiy bosqichlaridir. Bu situatsiya lingvistik tadqiqotlar uchun zamonaviy va qulay vosita hisoblangan korpus texnologiyalarida voqealanadi. Lekin hozirgi davrda internetdagi turli aloqa messengerlari alaqachon keng ko‘lamli infomaydonga aylangani rost. Binobarin, mazkur faktorning milliy boylik hisoblangan korpuslar yoki insoniyatga ko‘plab texnologik imkoniyatlar taklif qilayotgan AI dasturlari bilan integratsiyasi xoh maishiy hayot bo‘lsin, xoh sanoat, ilm-fan, savdo, ijtimoiy

⁵⁴ Niyazmetova K., Kuriyozov E. - O‘zbek tilidagi matnlarni sentiment tahlil qilish.

⁵⁵ Zaxarov V., Pavlovich, Mengliyev B., Hamroyeva Sh. Korpus lingvistikasi. – Toshkent, 2021. – 127-B.

⁵⁶ Захаров В.П., Богданова С.Ю. Корпусная лингвистика. Иркутск: ИГЛУ, 2011. - С.39.

⁵⁷ Xolmonova Zulxumor. Kompyuter lingvistikasi. – Toshkent, 2019. – 170-B.

tadqiqotlar kabi bir qancha sohalarga bevosita yoki bilvosita manfaat keltirishi mumkin. Ijtimoiy tarmoq matnlari esa quyidagicha sentiment tahlil qilinadi:

1. Ma'lumotlarni to'plash;
2. Preprocessing – matnni tozalash (hashtaglar, punktuatsion belgilar, emojilar, @mentionlardan tozalanadi, katta harflar kichik harflarga aylantiriladi)
3. Matn vektorlashtiriladi;
4. Model tanlanadi;
5. Matn tahlil qilinadi. (“positive”, “negative”, “neutral”)

Agar bitta gap tahlil qilinsa, odatiy tarzda baholanayveradi, lekin matnlarda holat boshqacharoq, ya'ni undagi gaplar sentiment xususiyatiga ko'ra guruhlariga ajratiladi. Qaysi guruhda kvantativ ustunlik bo'lsa, ana o'sha matnning sentiment labelini hal qiladi.

Keling, bir ijtimoiy tarmoqdagi fikrlarni tahlil qilib ko'raylik: (ushbu fikrlar Youtube platformasidagi “Texnoplov” kanalida e'lon qilingan “Siz radiatsiyaga chalingansiz – 15 DAQIQADA isbotlayman!” deb nomlangan videokommentlardan olingan)

Fikrlar:

Username: Doimgidek eng maksimal sifat bilan foydali kontent chiqaribsizlar... (ijobiy)

Username: Har doimgidek juda manfaatli va foydali. (ijobiy)

Username: Hayot borligi taxmin qilinayotgan sayyoralar haqida bo'lsin keyingi video... (neytral)

Username: Gipnozlar haqida bo'lsin keyingisi. (yurakchali emoji bilan) (ijobiy)

Yuqorida keltirilgan misollarda:

Neytral – 1 ta, negativ – 0, ijobiy – 3 ta. Demak, videoga bo'lgan fikr ijobiy.

Umuman olganda, modellar ishlab chiqilgan. Umumiy prinsiplar aniq, samarador. Bu borada, ijtimoiy tarmoq matnlarini sentiment tahlil qilishda, eng katta muammo – infarmatsion qamrovdagi obyektning son jihatdan ko'pligi va tartibsiz, keraksiz unsurlarga boyligi. Asosan, bu singari massenjerlararo tahlillar onlayn savdoda xaridor va sotuvchi munosabatlarini yanada unumdor qilishni ko'zlaydi, shuning uchun lo'ndalik hamda aniqlik eng avvalo erishilishi kerak bo'lgan omillardandir. Bundan tashqari mazkur tahlilning boshqa foydali jihatlari ham talaygina:

1. **Aholi salomatligi:** hissiyotlarni aniqlashning avtomatik usullari depressiyani aniqlashda foydalidir;

2. **Kiber zo'ravonlik holatlarini aniqlash va farovonlikni kuzatish;**

3. **Siyosat:** siyosatchilarga, saylov masalalariga, shuningdek, milliy va xalqaro voqealarga nisbatan jamoatchilikning, ayniqsa ijtimoiy tarmoqlardagi kayfiyatni kuzatishga qiziqish katta;

4. **Brend menejmenti, mijozlar bilan munosabatlarni boshqarish va fond bozori:** Bloglar, Facebook, va Twitter postlarining hissiyot tahlili allaqachon brend imidjini shakllantirish, mijozlar munosabatini kuzatish, mijozlar so'rovlari ustida ishlash va shikoyatlarini ko'rib chiqish uchun avtomatik dialog tizimlarini ishlab chiqishda keng qo'llaniladi;

5. **Ta'lim:** repetitorlik kurslari va talabalarni avtomatik baholash tizimlari javoblarning to'g'riligini aniqlash, shuningdek, ishtirokchining hissiy holatini aniqlash uchun javoblardagi his-tuyg'ularni aniqlaydi;

6. **Jamoatchilikni kuzatish:** Yuqorida aytib o'tilganidek, brend boshqaruvi va sog'liqni saqlash sohasida ishlashdan tashqari, yaqinda olib borilgan ba'zi ishlar hissiy ma'lumotlarning ijtimoiy tarmoqda qanday tarqalishini yaxshiroq tushunishga yordam bermoqda;

7. **Shaxsiy xususiyatlarni aniqlash:** odamlarning his-tuyg'ularini qanday ifodalashdagi belgilari kirishuvchanlik va o'ziga bino qo'yganlik kabi shaxsiy xususiyatlarning asosiy ko'rsatkichidir. Shunday qilib, yozma matndan shaxsiy xususiyatlarni aniqlaydigan ko'plab avtomatik tizimlar hissiyotlarni avtomatik aniqlashga tayanadi;

8. **Gender farqlarini aniqlash:** Erkaklar va ayollar ijtimoiy hayotda, ishda va hatto kompyuter orqali muloqotda turli lingvistik birliklardan foydalanadilar. Bir nechta tadqiqotlar

ushbu kontekstlarda erkaklar va ayollar tomonidan ishlatiladigan nutqidagi o'ziga xos his-tuyg'u o'rtasidagi farqni tahlil qilishgan;

9. Adabiy tahlil: Badiiy matnlarning katta to'plamlarini tahlil qilish uchun tabiiy tilni avtomatik qayta ishlash usullaridan foydalanishga qiziqish ortib bormoqda. Ayniqsa, his-tuyg'ularga nisbatan romanlar, pyesalar va kino ssenariylarida hissiyotlar oqimini kuzatish, katta matnlar to'plamiga xos bo'lgan his-tuyg'ularni aniqlash va syujet qahramonlarining his-tuyg'ularini kuzatish bo'yicha ishlar olib borilgan;

10. Lingvistik ekspertiza: Ushbu turdagi affektiv holatlarning har biri uchun hissiyotlarni ajratish orqali kimningdir blog postlari yoki tvitlari, yozishmalari depressiyadan dalolat beradimi yoki yo'qligini aniqlash mumkin⁵⁸.

Yuqorida sanab o'tilgan yo'nalishlar NLP texnologiyalari bilan qanchalik yaxshi integratsiya qilinsa, xususan, sentiment tahlil nazariyalari amaliyotga qanchalik faol joriy qilinsa, shu qadar jamiyatimiz uchun foydali bo'ladi. Turli xil sektorlarda mavjud muammolar: kadrlar, vaqt, qolaversa mablag' masalalaridagi takroriy kamchiliklar o'z yechimini topadi. Darhaqiqat, elektron texnologiyalardagi tezlik xususiyati o'zining foydalanuvchisiga ayni davrdagi sinxron ma'lumotlarni bera olishi bilan ham e'tiborga molikdir.

Foydalanilgan adabiyotlar:

1. Niyazmetova K., Kuriyozov E. – O'zbek tilidagi matnlarni sentiment tahlil qilish. (tezis) – 3-b.
2. Abdurahmonova N. Korpus lingvistikasi: korpus tuzish va undan foydalanish. – Toshkent, 2021.
3. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. Иркутск: ИГЛУ, 2011. – С.39.
4. Xolmonova Zulxumor. Kompyuter lingvistikasi. – Toshkent, 2019. – 247-b.
5. Maqolada AI dasturlarining dasturiy prinsiplari haqidagi savollarga berilgan javobdan ham foydalanildi.
6. Swati Redhu, Sangeet Srivastava, Barkha Bansal, Gaurav Gupta. Sentiment Analysis Using Text Mining: A Review. // International Journal on Data Science and Technology. – 5-b.
7. Allanazarova S, Elova D. Sentiment tahlil uchun matndan valentlik, hissiyot va ta'sirchan holatlarni aniqlash. // O'zbekiston til va madaniyat jurnali. – Toshkent, 2024

⁵⁸ Allanazarova S, Elova D. Sentiment tahlil uchun matndan valentlik, hissiyot va ta'sirchan holatlarni aniqlash. // O'zbekiston til va madaniyat jurnali. – Toshkent, 2024. – 38-B.

O'ZBEK ADABIY TILINING QOZOQ TILIGA TA'SIRI

Jalg'asbekova Madina Qaldibay qizi

Aniq va Ijtimoiy fanlar universiteti
O'zbek lingvistikasi yo'nalishi 1-kurs magistranti

Annotatsiya: Ushbu tezisdan o'zbek tilining qozoq tiliga bo'lgan ta'sirining tarixiy, lingvistik va madaniy asoslari tahlil qilinadi. O'zbek va qozoq xalqlari o'rta asrlardan boshlab Markaziy Osiyoning turli hududlarida qo'shnichilik, savdo-sotiq, harbiy va siyosiy aloqalar orqali bir-biriga yaqinlashgani haqida so'z boradi. Xususan, Temuriylar davrida Movarounnahr madaniyati gullab-yashnagan bir davr bo'lib, bu davrda o'zbek adabiy tili (eski o'zbek tili yoki chig'atoy tili) butun Markaziy Osiyoda, shu jumladan, hozirgi Qozog'iston hududida ham keng tarqalganligi haqida ma'lumotlar keltiriladi.

Kalit so'zlar: lingvistik ta'sir, undoshlar, sintaktik daraja, morfologik jihat, madaniy til

Abstract: In this thesis, the historical influence of the Uzbek language on the Kazakh language, linguistic and cultural foundations are analyzed. It is said that the Uzbek and Kazakh peoples have come closer to each other in various regions of Central Asia since the Middle Ages through neighborliness, trade, military and political relations. In particular, during the Timurid period, the Movarounnahr culture flourished, and during this period, the Uzbek literary language (the old Uzbek or Chigatoi language) was widespread throughout Central Asia, including the territory of present-day Kazakhstan. information is provided.

Keywords: linguistic influence, consonants, syntactic level, morphological aspect, cultural language.

Абстрактный. В данной диссертации рассматривается историческое влияние узбекского языка на казахский язык. Анализируются лингвистические и культурные основы. Говорят, что узбекский и казахский народы сблизились друг с другом в различных регионах Центральной Азии со времен Средневековья посредством добрососедства, торговли, военных и политических отношений. В частности, в период Тимуридов произошел расцвет культуры Моваруннахр, и в этот период узбекский литературный язык (староузбекский или чигатойский язык) получил широкое распространение по всей Средней Азии, включая территорию современного Казахстана. информация предоставлена.

Ключевые слова: языковое влияние, согласные, синтаксический уровень, морфологический аспект, культурный язык

Kirish

Turkiy tillar oilasiga mansub bo'lgan o'zbek va qozoq tillari asrlar davomida bir-biriga yaqin geografik, ijtimoiy, siyosiy va madaniy muhitda rivojlanib kelgan. Bu tillar o'zaro aloqadorlik, o'xshashlik va o'zaro ta'sir jihatdan juda boy tarixga ega. Ayniqsa, o'zbek tilining qozoq tiliga ko'rsatgan ta'siri tilshunoslar e'tiboridan chetda qolmagan. Bu ta'sir faqat leksik darajada emas, balki fonetik, morfologik va sintaktik darajalarda ham seziladi. Chig'atoy tili, ya'ni eski o'zbek adabiy tili, Abdurahmon Jomiy, Alisher Navoiy kabi buyuk adiblar tomonidan taraqqiy ettirildi va bu til mintaqaning madaniy tiliga aylandi. Ko'plab qozoq shoirlari va yozuvchilari chig'atoy tilida ijod qilgan yoki bu tilning ta'sirida bo'lgan. O'zbek tilining qozoq tiliga ta'sirining bir necha jihatlarini quyida ko'rib chiqamiz.

Leksik ta'sir. O'zbek tilining qozoq tiliga eng sezilarli ta'siri so'z boyligi darajasida bo'lgan. Ayniqsa, ma'naviy, diniy va madaniy sohalariga oid so'zlar o'zbek tilidan qozoq tiliga ko'chgan. Misol:

- Kitap, mektep, bilim, ziyatker, ustaz kabi so'zlar o'zbek tilidan olingan bo'lib, hozirgi qozoq tilida keng ishlatiladi.

- Ko'plab arabcha va forsha ildizli so'zlar o'zbek tilidan o'tgan ko'rinishda qozoq tiliga kirib kelgan. Masalan, adabiet (adabiyot), dasturxon (dasturxon), sauap (savob) kabi so'zlar.

Fonetik va fonologik o'xshashlik va ta'sirlar. O'zbek va qozoq tillari fonetik tizimi o'xshash bo'lsa-da, ayrim farqlar mavjuddir. Shunga qaramay, ba'zi o'zbekcha talaffuz unsurlarining qozoq tilida paydo bo'lishi, ayniqsa janubiy qozoq lahjalarida, o'zbek tilining fonetik ta'siridan dalolat beradi. Masalan, "q" tovushi o'rniga "k" tovushining ishlatilishi, yoki "g" tovushi o'rniga "g" tovushining tanlanishi bu hududlarda kuzatiladi. O'zbek tili — Oltoy tillari oilasining turkiy tillar turkumiga kiruvchi tildir. O'zbek adabiy tili tarixi quyidagi davrlarni o'z ichiga oladi: Miloddan oldingi davrlardan to X asrlargacha bo'lgan davr. Bu davrdagi til fanda qadimgi turkiy til deb yuritiladi. Qadimgi xalq og'zaki ijodi namunalari, O'rxun-Enasoy yodgorliklari (VI— VII asrlar) shu tilda yaratilgan. Qozoq xalqining ona tili, Qozog'iston Respublikasining davlat tili. G'arbiy turkiy tillarning qipchoq guruhiga kiruvchi qozoq tili arab-fors tillaridan ko'p so'z o'zlashtirmagan, turkiy til kategoriyalarini o'zida chuqur saqlaydigan, shuningdek, dialektal xususiyatlari zaif monolit tillardan biridir. Ma'lumki, o'zbek tilida qattiq undosh k, q tovushlari unli bilan boshlangan qo'shimchalar bilan qo'shib, so'z oxirida qattiqlashadi, so'z oxirida esa- g, g' tovushlari paydo bo'ladi. Bu o'zgarish yozuvda ham, talaffuzda ham hisobga olinadi. Masalan: tilak+ing=tilaging, yurak+im=yuragim, kubok+i=kubogi, bek+i=begi; tayoq+i=tayog'i, qoshiq+i=qoshig'i kabi. Lekin ko'p bo'g'inli o'zlashma so'zlarga va bir bo'g'inli ba'zi so'zlarga egalik qo'shimchasi qo'shilganda k,q tovushi aslida aytiladi va yoziladi: ishtirok - ishtiroki, erk - erki, huquq - huquqim, ravnaq - ravnaqi kabi. Qozoq tilida, o'zbek tilida bo'lgani kabi, undoshlar ham repressiv ovozga ega. k, k, n tovushlari unli bilan boshlangan qo'shimchalar bilan qo'shib, so'z oxirida qattiqlashadi, so'z oxirida esa- f, r, b (ba'zan y) tovushlari paydo bo'ladi. Bu o'zgarish yozuvda ham, talaffuzda ham hisobga olinadi.

Masalan: тарақ+ым=тарағым, күрек+i=күрегі, дон+ы=добы va hokazo.

Turkiy tillarning asl tabiatining bu ustun belgilari, turkiylar hayotidagi lingvokontaktologik jarayonlar natijasida hosil bo'lgan fonetik, morfologik, leksik xususiyatlar, masalan, turkiy tillarning avlod tuzilishida ham o'z izini qoldirgan. Turkiy tillar bo'g'in tuzilishiga ko'ra to'rt guruhga bo'linadi:

V- bir unli bilan yasalgan bir bo'g'in:

CV-unli bilan tugaydigan ochiq bo'g'in:

VC, VCC - unlilardan undoshlarga otiladigan yopiq bog'inlar:

CVC, CVCC - undosh bilan boshlanib, undosh bilan tugagan yopiq bo'g'in: O'zbek va qozoq alifbolarini solishtirish. Bu ikki turkiy tilning alifbosini tuzilishi, asosan, arab alifbosidan kelib chiqqan. O'zbek va qozoq tilining o'xshash jihati shundaki, ikki tilda ham lotin, ham krill alifbolari mavjud.

Morfologik ta'sir. Morfologik tahlilni avtomatlashtirish tabiiy tilning so'z tuzilmalari va so'z yasallishining rasmiy tuzilmalarini aniqlash, shuningdek, uning strukturaviy nazariyasi va kompyuter modelini yaratish bilan bog'liq masalalarni hal qilish imkonini beradi. Chunki morfologik tahlil axborotni izlash, bir tildan ikkinchi tilga mashina tarjimasini, savollarga semantik javoblar, avtomatik qaror qabul qilish, xulosalarni tekshirish kabi intellektual tizimlarni yaratishda foydalaniladi. O'zbek tili oltoy tillari oilasining turkiy tillar turkumiga [J.I.Xolmurodov O'zbek tili 2019,173-b.], qozoq tili esa turkiy tillarning qipchoqlar guruhiga kiradi [Qozoq tili grammatikasi 2002.784-bet]. Lekin ular umumiy turkiy tillarga xos aglyutinatativ xususiyatga ega. Bu xususiyat har bir o'zak so'zga affikslarni (qo'shimchalar va bog'lanishlar) birlashtirib, yangi so'z hosil qilish yoki turli so'z yasash qobiliyati bilan tavsiflanadi. Bu yerda qo'shimchalar so'zning semantikasini (ma'nosini) o'zgartirib, yangi so'z hosil qiluvchi semantik turkumga, qo'shimchalar esa so'zning faqat tuzilishini o'zgartiruvchi tuzilma turkumiga kiradi. Turkiy tillarning bu xususiyati ularning morfologik qoidalarini osongina rasmiylashtirish imkonini beradi. Turkiy tillarning morfologik pozitsiyalarini rasmiylashtirish va morfologik tahlil qilish bo'yicha ko'plab tadqiqotlar o'tkazildi. O'zbek tili uchun morfologik tahlil ishlari uzoq vaqt davom etgan va bu bo'yicha juda samarali

ishlar olib borilgan va olib borilmoqda. O‘zbek va qozoq tillaridagi har bir so‘z turkumining o‘zgarish tizimini ta‘minlovchi o‘ziga xos so‘z yasovchi qo‘shimchalar va birikmalarga ega. O‘zbek tilidagi ayrim so‘z yasovchi va grammatik affikslarning qozoq tilida ham paydo bo‘lishi yoki o‘xshash funksiyani bajarishi kuzatilgan.

Sintaktik o‘xshashliklar va farqlar. Har ikki til ergativ strukturaga ega emas, ammo turkiy tillarga xos SOV (Subyekt–Obyekt–Fe‘l) tartibini saqlaydi. Shunga qaramay, o‘zbek tilidagi ba‘zi gap tuzilmalari, ayniqsa adabiy uslubda, qozoq tilida ham o‘z aksini topgan.

Madaniy va adabiy ta‘sir. O‘zbek adabiyoti va madaniyati qozoq xalqining ma‘naviy hayotiga ham kuchli ta‘sir ko‘rsatgan. Alisher Navoiy, Zahiriddin Muhammad Bobur, Abdurauf Fitrat, Abdulla Qodiriy kabi adiblarning asarlari qozoq ziyoilari tomonidan o‘rganilgan va tarjima qilingan.

Hududiy yaqinlik va ijtimoiy aloqa orqali ta‘sir. Janubiy Qozog‘iston viloyatlari aholisi orasida o‘zbeklar bilan qozoqlar doimiy aloqada yashagani sababli kundalik hayotda til aralashuvi kuchli bo‘lgan. Ayniqsa bozor, oila, qo‘shnichilik, to‘y-marosimlar kabi ijtimoiy muhitda ikki til bir-biriga kuchli ta‘sir ko‘rsatgan.

Xulosa

Xulosa qilib aytganda, O‘zbek tilining qozoq tiliga bo‘lgan ta‘siri asosan leksik, fonetik va madaniy darajalarda seziladi. Bu ta‘sir tarixiy asoslar, madaniy aloqalar va hududiy yaqinlik orqali shakllangan. Bugungi kunda ham ikki tilda yashovchi hududlarda bu ta‘sir davom etmoqda.

Foydalanilgan adabiyotlar ro‘yxati:

1. J.I.Xolmurodov O‘zbek tili . Toshkent :,2019,173-b
2. Qozoq tili grammatikasi . Ostona:,2002.784-bet
3. Mahmudov N. va boshqalar. O‘zbek tili. – Toshkent: O‘qituvchi, 2012.
4. Ziyoyev R. Qozog‘iston o‘zbeklari tilining leksik xususiyatlari. – Toshkent: Fan, 2004
5. Isayev M. Qozog‘iston til siyosati va o‘zbeklar. – Almaty: Arys, 2011.
6. Kononov S. Grammatika sovremennogo uzbekskogo literaturnogo yazyka. – Moskva: Nauka, 1960.
7. Shoniyozov K. O‘zbek va qozoq xalqlarining etnik aloqalari. – Toshkent: Fan, 1990.
8. Muhammadjonov A. Turkiy tillar o‘zaro ta‘siri haqida. – “Til va adabiyot” jurnali, 2017-yil, 2-son.

O'ZBEK TILINING SIFAT LEKSEMALARIDA KONTEKSTOLOGIK SINONIMLAR BAZASINI YARATISH MASALALARI

Mohichehra Abramatova Abdurazzoq qizi,
Toshkent davlat Sharqshunoslik universiteti
1-kurs magistranti
E-mail: abramatovamohichehra@gmail.com

Ilmiy rahbar: **Abduraxmonova Nilufar Zaynobiddin qizi**
O'zMU professori, filologiya fanlari doktori

Annotatsiya: Matnlarni semantik talqin qilishda soʻz ma'nolarini tushunish markaziy oʻrinni egallaydi. Sinonimik qatordagi muqobillarni tanlash masalasi NLP sohasining ma'lumot olish, mashina tarjimai, sentiment tahlil va boshqa sohalar uchun muhim sanaladi. Ushbu maqolada korpus tahlil asosida soʻz ma'nolarining kontekstologik sinonimlar bazasini yaratish va ularni semantik tahlilda ifodalash masalalariga doir ayrim nazariy qarashlar tahlilga tortilgan. Shuningdek, kuzatishlarimizda mavjud sinonimlar lugʻatlarini tahlil qilgan holda oʻzbek tili korpusi uchun sinonimik birliklar tizimini yanada takomillashtirishda korpus tahlilining muayyan afzalliklari haqida batafsil ma'lumotlar keltirilgan.

Kalit soʻzlar: Kompyuter, lingvistika, korpus, konteks, sinonim, sifat, leksema, baza, NLP.

Annotation: Understanding the meaning of words occupies a central place in the semantic interpretation of texts. The issue of choosing alternatives in a synonymous line is considered important for information retrieval, machine translation, sentiment analysis and other fields of NLP. In this article, some theoretical views on the issues of creating a database of contextual synonyms of word meanings based on corpus analysis and expressing them in semantic analysis are presented. drawn. Also, our observations provide detailed information about specific advantages of corpus analysis in further improving the system of synonymous units for the corpus of the Uzbek language by analyzing existing dictionaries of synonyms.

Keywords: Computer, linguistics, corpus, context, synonym, adjective, lexeme, base, NLP.

Аннотация: Понимание значения слов занимает центральное место в смысловой интерпретации текстов. Вопрос выбора альтернатив в синонимическом ряду считается важным для поиска информации, машинного перевода, анализа настроений и других областей НЛП. В данной статье представлены некоторые теоретические взгляды на вопросы создания базы данных контекстных синонимов значений слов на основе корпуса. анализ и выражение их в семантическом анализе. Также наши наблюдения дают подробную информацию о конкретных преимуществах корпусного анализа в дальнейшем совершенствовании системы синонимических единиц корпуса узбекского языка путем анализа существующих словарей синонимов.

Ключевые слова: Компьютер, лингвистика, корпус, контекст, синоним, прилагательное, лексема, основа, NLP.

Ma'lumki, bir soʻz qoʻshimcha ravishda son-sanoqsiz ma'nolarni va munosabatlarni ifodalashi mumkin. Muayyan soʻz asosiy "lugʻat" ma'nosiga koʻra koʻpincha bir-biridan farq qiladigan yaqin sinonimlarga ega va undan faqat shu ma'noda va oʻsha konteksdagina qoʻllana olishi mumkin. Toʻgʻri soʻzni tanlay olish uchun har qanday muayyan vaziyatda barcha variantlar orasidagi farqlarni diqqat bilan koʻrib chiqish kerak. Muqobil soʻzni tanlash odamlar uchun qiyin

bo'lishi mumkin, ammo hozirgi kompyuter tizimlari va bu funksiyani bajara oladigan shunday dasturlar yaratilyabdiki, bu borada yordam bera oladi.

Sinonimlar - bir umumiy ma'noga ega bo'lgan, denotativ ma'nosi bir xil, qo'shimcha ma'nosi uslubiy va boshqa munosabatlarni ifodalovchi xususiyatlari bilan o'zaro farqlanadigan til birliklari - so'z, ibora va sintaktik birliklar sanaladi. Masalan: Chiroyli, go'zal, maftunkor, husnli, sohibjamol, zebo, suluv kabi so'zlar sinonim so'zlardir. Kontekstli sinonimlar esa o'xshash ma'noga ega bo'lgan, lekin muayyan kontekst uchun eng mos keladigan so'zlar hisoblanadi. Umuman olganda, bir-birining o'rnida qo'llanilishi mumkin bo'lgan standart sinonimlardan farqli o'laroq, kontekstli sinonimlar vaziyat, ohang yoki muhitning unsurlarini hisobga oladi⁵⁹. Misol uchun, "baxtli" so'zi quyidagi kabi kontekstli sinonimlarga ega bo'lishi mumkin: - "quvonchli" (ko'proq qizg'in, bayram konteksti) – "mamnun" (ko'proq bo'ysunuvchi, tinch kontekst) - "hursand" (juda hayajonlangan, ko'pincha ma'lum bir voqeadan keyin). Kontekstologik sinonimlar - bu so'zlarning yoki iboralarining kontekstdagi ma'nosi va ishlatilishini o'rganadigan lingvistik yo'nalish hisoblanadi. Dunyo bo'ylab bu mavzuda izlanish olib borayotgan bir qancha olimlar va ularning ishlari mavjud, jumladan, J. R. Firth, ingliz lingvisti, kontekstualizm va "kontekst" tushunchasini rivojlantirgan. U "ma'no kontekstdan kelib chiqadi" degan fikrni ilgari surgan, leksikografiya va tilni o'rganishda kontekstning ahamiyatini ta'kidlagan. Uning ishlari "lingvistik kontekst" tushunchasini rivojlantirishda muhim ahamiyat kasb etgan.

M.A.K.Halliday (Michael Alexander Kirkwood Halliday) zamonaviy lingvistikaning asosiy figurali va funksional tilshunoslik nazariyasining asoschisi sifatida tanilgan ingliz olimi tilning ijtimoiy funktsiyalarini va kontekstda sinonimlar qanday ishlatilishini o'rganishda muhim hissa qo'shgan. Uning "Systemic Functional Linguistics" asari tilning ijtimoiy va kontekstual jihatlarini o'rganishga qaratilgan.

W.Croft til va uning kontekstual ishlatilishini o'rganib, lingvistik struktura va sinonimlar orasidagi bog'lanishlarni ta'kidlagan. "Sistematika va Semantika" va "Typology and Universals" kabi asarlarda kontekstual sinonimlar haqida yozgan.

A.P.Cowie ingliz lingvisti bo'lib, asosan leksikografiya, semantika va korpus lingvistikasi sohalarida ishlagan. U leksik tizimlar, sinonimlar va ularning kontekstual ishlatilishi bo'yicha ko'plab tadqiqotlar o'tkazgan. "Oxford Dictionary of English Idioms" asarida u kontekstga bog'liq sinonimlar haqidagi ma'lumotlarni jamlagan. Olim leksikografiya va semantika sohalarida kontekstual sinonimlarni o'rganishga bag'ishlangan ko'plab maqolalar ham yozgan.

Z.D.Radford kontekstual sinonimlar va ularning grammatik funktsiyalarini o'rganishga e'tibor qaratgan. Uning ishlari kontekstda sinonimlarning qanday ishlatilishini va ularning o'zaro aloqalarini ko'rsatadi. "Syntax: A Generative Introduction" (1988) asari uning sintaksisga oid muhim ishlaridan biridir, u tilni yaratuvchiligini va uni qanday tushunish kerakligini ta'kidlagan.

J.Sinclair ingliz tilshunosi, leksikografiya sohalarida katta hissa qo'shgan va zamonaviy tilshunoslikda korpus tadqiqotlarini rivojlantirishga yordam bergan. "Collins Cobuild English Dictionary" ni yaratishda ishtirok etgan. Uning ishlarida kontekstual sinonimlar va kolokatsiyalar o'rtasidagi aloqalar o'rganilgan. "Collocations Dictionary" asarida so'zlarning birgalikda ishlatilishi va ularning kontekstdagi ahamiyatini tahlil qilib bergan.

T.Givon zamonaviy lingvistikada taniqli olim bo'lib, asosan semantika, sintaksis va pragmatika sohalarida muhim ishlar olib borgan. U tilning tuzilishi va ishlatilishi o'rtasidagi bog'lanishlarni o'rganishga qaratilgan yondashuvlari bilan tanilgan, "Syntax: A Functional-Typological Introduction" asarida tilning kontekstual jihatlari haqida fikrlar bildirgan, tilning kontekstga bog'liqligini va sinonimlar orasidagi nozik farqlarni o'rganishda muhim hissa qo'shgan.

Tabiiy til avlodida leksik tanlashning eng kam tushunilgan jihatlardan biri bu yaqin sinonimlar orasidan tanlashdir. Edmonds va Xirst kabi olimlar bu masala bo'yicha oldingi tadqiqotlarda leksikograflar tomonidan tahlil qilinganidek, yaqin sinonimlar o'rtasidagi semantik farqlarga e'tibor qaratgan. Ob-havo prognozlarida sinonimga yaqin tanlovni empirik tahlil qilish

⁵⁹ <https://bestpublication.org/index.php/pedg/article/download/8950/9087/8359>

shuni ko'rsatadiki, boshqa omillar semantik farqlardan ko'ra muhimroqdir. Bularga individual mualliflarning afzalliklari va o'ziga xos xususiyatlari kiradi. Bu tahlil shuni ko'rsatadiki, semantik farqlar sinonimlarga yaqin tanlovga ta'sir qilganda, ular buni muallifga bog'liq holda qilishlari mumkin. Avvalo, Natural Language Generation (NLG) tizimlari yaratilgan matnda ishlatiladigan so'zlarni tanlashi kerak. Bu vazifa leksik tanlash deb ataladi. Agar NLG tizimi bir nechta sinonimlar yoki yaqin sinonimlar bilan ifodalanishi mumkin bo'lgan ma'noni yetkazishni xohlasa, leksik tanlash, ayniqsa, qiyin vazifa bo'lishi mumkin. Misol uchun, agar tizim 12:00 vaqtini bildirmoqchi bo'lsa, uni peshin, tush, erta peshin deb ifodalashi kerakmi? Bu nuqtai nazardan leksik tanlashning ikki jihati bor: izlanish va tanlash. Qidiruv - kerakli ma'lumotni bildiruvchi leksemalarni yoki leksemalar turkumini topish vazifasi, tanlov - bu mavjud sinonimlarning qaysi birini haqiqatda ishlatishni tanlash vazifasi. Bu erda bizning e'tiborimiz tanlovga, xususan, yaqin sinonimlar orasidagi tanlovga qaratilgan. Ehtimol, yaqin sinonimlar o'rtasida tanlov bo'yicha eng mashhur oldingi ish Edmonds va Hirstdir. Ular so'zlar yaqin sinonimlarning "klasterlariga" guruhlangan deb taxmin qilishadi. Har bir klaster asosiy ma'noga yoki denotatsiyaga ega bo'lib, klasterdagi barcha so'zlar o'zaro bog'lanadi. Klasterdagi yaqin sinonimlar asosiy ma'noga mos keladigan nozik ma'no yoki konnotatsiyani ko'rsatadigan keyingi "periferik" farqlar bilan ajralib turadi. Bu farqlar standart sinonimik lug'atga asoslanadi. Haqiqiy leksik tanlash moslashuv va optimallashtirish jarayoniga asoslanadi, u hozirgi vaziyatga eng yaqin sinonimni qidiradi. Edmonds va Xirstning ishi ta'sirchan, ammo ular o'zlari korpus tahlili yoki psixolingvistik tajribalar o'tkazmaganligi uchun empirik asoslanmagan (garchi ular ishlatgan sinonim lug'atdagi ta'riflar ba'zi empirik asosga ega bo'lsa ham)⁶⁰.

C.Baclayon and R.Roxas "Designing a Context-Based English Synonym Database" maqolasida kontekstga asoslangan sinonimlar ma'lumotlar bazasini taqdim etadi, bu inson yoki mashinaga so'zlarning tegishli sinonimlarini jumlada ishlatilishiga qarab topishga yordam beradi. Ma'lumotlar bazasi WordNet-ga o'xshash bo'lib, muhim ma'lumotlar uchun qo'shimcha maydonlarga ega: bo'g'inlar soni, oxirgi bo'g'inning talaffuzi va mumkin bo'lgan kontekst so'zlari. Ma'lumotlar SQLite va Java orqali kirish mumkin bo'lgan relyatsion ma'lumotlar bazasida saqlanadi. Qo'lda baholash shuni ko'rsatadiki, ma'lumotlar bazasi 79,81% o'rtacha aniqlik bilan to'g'ri sinonimlarni qaytarishga qodir.⁶¹

Rus olimlaridan A.A.Potebnya til va fikr o'rtasidagi munosabatni o'rganib, sinonimlarning ma'nosi kontekstga bog'liq ekanligini ta'kidlagan. U sinonimlar o'rtasidagi farqlar ko'pincha ularning foydalanish kontekstida yashiringanligini ko'rsatgan, sinonimlarning semantik xususiyatlarini, ularning o'zaro bog'liqligini va qanday qilib biror kontekstda turlicha ma'no kasb etishini va bu jarayonda, sinonimlar orasidagi nozik farqlarni aniqlashga harakat qilgan, tilning fikrni ifodalashdagi rolini o'rganish jarayonida, sinonimlar orqali o'z fikrlarimizni qanday ifodalashimiz mumkinligini, bundan tashqari sinonimlarning madaniy va mifologik kontekstdagi ahamiyatini o'rganib, til va madaniyat o'rtasidagi bog'liqlikni tahlil qilgan. Yana bir rus tilshunosi V.Vinogradov kontekstologik sinonimlar mavzusida bir necha muhim asarlar yozgan va uning ishlari asosan stilistika, semantika va sinonimlar o'rtasidagi o'zaro munosabatlarga qaratilgan. U o'zining "Russkaya leksikologiya" asarida rus tilidagi sinonimlar, ularning stilistik va semantik xususiyatlari haqida chuqur tahlillar olib borgan, "O sinonimakh" maqolasida sinonimlar va ularning kontekstdagi rolini, shuningdek, sinonimlar o'rtasidagi nozik farqlarni o'rgangan, "Stylistics of the Russian Language"da rus tilining stilistik jihatlari, shuningdek, sinonimlarning o'zaro aloqalari va kontekstda qanday ishlatilishi haqida fikr yuritgan. O.A.Vasilieva esa sinonimlarning turli kontekstlarda qanday ishlatilishini va ularning muloqotdagi rolini o'rgangan, semantik analiz, stilistika va leksikologiya, yani badiiy va akademik matnlardagi sinonimlarning stilistik jihatlari ko'rib chiqqan. Bundan tashqari A.P.Cherepanov - sinonimlarning semantik jihatlari va kontekstga bog'liqligini, Y.A. Kuznetsov - sinonimlarning leksik va stilistik

⁶⁰ https://www.academia.edu/109694163/Semantic_Representations_of_Near_Synonyms_for_Automatic_Lexical_Choice

⁶¹ https://www.researchgate.net/publication/320675722_Designing_a_Context-Based_English_Synonym_Database

xususiyatlarini, D.N.Sokolov - sinonimlarning stilistik funktsiyalarini va ularning mulohazadagi rolini o'rganish.

O'zbekistonda ham bu sohada ilmiy izlanishlar haligacha davom etmoqda. Kontekstologik sinonimlar bo'yicha ilmiy tadqiqotlar olib borayotgan olimlar o'z asarlarida tilni yanada chuqurroq tushunishga, o'zbek tilining leksik va sintaktik boyligini oshirishga e'tibor qaratishmoqda va ularning ishlari o'zbek tilshunosligi va leksikografiya sohalarida muhim ahamiyatga egadir. Jumladan, filologiya fanlari doktori, professor, O'zbekiston Respublikasi Fanlar akademiyasi akademigi Xodjiyev Azim Po'latovichning O'zbek tili morfologiyasi, so'z yasalishi, leksikologiya va leksikografiya, orfografiyaga oid bir qator ishlari o'zbek tilshunosligi rivojiga munosib hissa qo'shdi. "O'zbek tilida ko'makchi fe'l", "To'liqsiz fe'l", "Hozirgi o'zbek tilida shakl yasalishi", "O'zbek tilining izohli lug'ati sinonimlari", "O'zbek tilshunoslik atamalarining izohli lug'ati", "Lingvistik atamalarining izohli lug'ati" asarlari shular jumlasidandir. Filologiya fanlari doktori, professor Shavkat Raxmatullayev Ubaydullayevich ham o'zbek tili leksikologiyasi, leksikografiyasi, morfologiyasi rivojiga ulkan hissa qoshgan olimdir. Xususan, o'zbek frazeologiyasining shakllanishi va rivojlanishida uning xizmatlari katta. "O'zbek tilining izohli frazeologik lug'ati", "O'zbek tili antonimlarining izohli lug'ati", "O'zbek tilining izohli lug'ati" shular jumlasidandir. Olim "O'zbek tilining izohli lug'ati"da sinonim, sinonimik, sinonimiya terminlariga quyidagicha ta'rif berilgan: Sinonim (yunoncha- bir xil nomli) tilshunoslikning shakli bir xil, ma'nosi esa bir xil yoki biroz farqli bo'lgan til birliklari qatoridagi til birligi. Sinonimik - tilshunoslikda bir-biriga sinonim bo'lgan ma'nodosh, sinonimik so'zlar. Sinonimika tilshunoslikda:

1. Biror tilning sinonimlari majmui. O'zbek tili sinonimikasi.
2. Leksikologiyaning sinonimlarni o'rganuvchi bo'limi.

Sinonimiya - til birliklarining ma'no jihatidan yaqinlik, ma'nodoshlik hodisasi. H.Jamolxonov "Hozirgi o'zbek adabiy tili" kitobining "Leksik sinonimiya" qismida "Sinonimlar nutqda muhim uslubiy vosita sanaladi, ular nutqning ravon va ta'sirli bo'lishini, fikrning aniq va obrazli chiqishini ta'minlaydi" degan fikrni berib o'tganlar. Sinonimiya - leksemalarning bir-xil ma'no anglatishiga ko'ra guruhlanishi. Sinonimik qator (sinonimik uya) – bir ma'no anglatishiga ko'ra guruhlangan sinonim so'zlar yig'indisi. Dominanta – sinonimik qatordagi betaraf ma'noli so'z. Ma'no sinonimlari – bir umumiy ma'noga ega bo'lsa-da, ma'no qirralari bilan farqlanuvchi sinonimlar. Uslubiy sinonimlar – leksik ma'nolari ijobiy yoki salbiy munosabat ifodalovchi semalar bilan qoplangan sinonimlar. Nutqiy sinonimlar – nutqning u yoki bu turiga xoslanganligi bilan farqlanadigan sinonimlar⁶².

Sinonimlarning turli xususiyatlarini o'rganish sohasida o'zbek tilshunosligida bir qator yutuqlarga erishilgan. Shunga qaramasdan akademik Azim Hojiyev sinonimlarning o'rganilishi tarixi haqida quyidagicha mulohaza bildiradi: "Sinonimlar turli tillarda har xil darajada o'rganilgan. Bu masala ayrim tillarda ancha to'la o'rganilgan bo'lsa, ayrim tillarda, jumladan, o'zbek tilshunosligida deyarli ishlanmagan. Tilshunoslikda bayon qilingan fikrlar bir-biriga ko'p hollarda mos kelmaydi. Bu holat, avvalo, sinonimlarning ta'rifida ko'zga tashlanadi". Bundan tashqari o'zbek olimlaridan F.A.Kichkineva, A.K.Sirojiddinov, A.K.Sirojiddinov, O.M.Tursunov, N.R.Nematova kabi olim va olimlar kontekstual sinonimlar va ularning leksik ma'nolari bo'yicha tadqiqotlar olib borgan, so'zlarning kontekstga bog'liq o'zgarishlarini o'rganish va bu borada ilmiy maqolalar chop etgan.

Umuman olganda, kontekstni tushunish muayyan sinonim qaysi leksikaga mos kelishini aniqlashga yordam beradi. Masalan, "abadiy" so'zining "mangu", "ilalabad", "toabad", "umrbot" kabi sinonimlari mavjudligini bilamiz, ammo yana bir qator so'zlar ham borki ma'lum kontekstagina ushbu so'z bilan sinonimlik qila oladi. Quyida ularning bir qanchasini misollari bilan ko'rib chiqamiz:

⁶² <https://in-academy.uz/index.php/cajei/article/download/23339/15401/22063>

“Ajablanarlisi shuki, kasal bo‘lib yotmadilar. Avval Er ketdi. Tungi uyquasi abadiy uyquning debochasi ekanini bilmay ketdi. Kechqurun yotdi-yu, tongda uyg‘onmadi.” (Tohir Malik, “Shaytanat 2- kitob”, 1991).

Bu kontekstda “abadiy uyqu” ga “o‘lmoq, vafot etmoq” so‘zlari sinonim bo‘la oladi;

“Bu mungli qissa shaytonsaroyning gunoh kitobiga abadiy yozilib qoldi.” (Tohir Malik, “Shaytanat 2- kitob”, 1991)

Bunda “abadiy” so‘zi “o‘chmas” so‘zi bilan sinonimlik hosil qiladi;

“Bedil mutolaasi Otaga abadiy haqiqatlardan, o‘lmas va yuksak tuyg‘ulardan saboq beradi. Abdulqodir Bedil she‘riyatida, jumladan, shunday misralarni: Maqomi zolim oxir bar zaifonast arzoniy, Agar otash furu aftad baxokistar dehad joro (Mazmuni: Zolimning maqomi - qudrati, zulmi oxir- oqibat zaiflashib, kuchdan qoladi.” (‘G‘afur G‘ulom “Sog‘inish”)

Bu kontekstda “abadiy” so‘zi “chin, asl” so‘zlari bilan sinonimlik hosil qiladi;

“Nima uchun u birdaniga hura ketadi? Hamma gap shundaki, yozuvchi mana shu mitti kuchukcha taqdirida odamzodga xos bo‘lgan muhim va abadiy muammolardan biriga javob izlagan. Ya‘ni dunyodagi har bir mavjudot, ayniqsa, inson, avvalo, o‘zini, o‘zligini tanimaguncha, kimning avlodi, asosiy vazifasi nimadan iborat ekanini anglamagunga qadar hayotda o‘z o‘rnini hech qachon topa olmaydi.” (Janni Rodari “Qutbdagi gunafsha”)

Bu matndagi “abadiy” so‘zi “qiyin, murakkab” so‘zlari bilan sinonimlik hosil qiladi. (abadiy muammo – qiyin muammo);

“Biroq shoir ijodiy tafakkuri butun insoniyat hayoti - uning quvonch-u tashvishlari, hatto qismati ham mana shu lahzalarga jo ekanini favqulodda misollar bilan isbotlab Bir onning bahosin o‘lchamoq uchun oltindan tarozu, olmosdan tosh hayotning abadiy qonuniyatlarini anglagan, his qilgan odam uchun behuda lahzaning o‘zi yo‘q va bo‘lishi ham mumkin emas. Zero, inson necha yil umr ko‘rishidan qat‘i nazar, bu umrning asosini mana shu lahzalar tashkil qiladi.” (G‘afur G‘ulom “Vaqt”)

Bu kontekstda “abadiy” so‘zi “adolatli, haqiqiy” so‘zlari bilan sinonimlik hosil qiladi;

“U o‘zining baxt-u iqboli yo‘lida kurashgan ulug‘lar bilan birga, odamzod tirikligiga xavf-u qo‘rqinch solgan yovuz kimsalarni ham unutmaydi. Ularning qilmishlari takrorlanishidan ogoh faylasuf shoir G‘afur G‘ulom o‘zining mo‘jazgina asari – “Vaqt” she‘ri bilan lahzaga abadiy haykal qo‘ydi, deb bemalol aytsak bo‘ladi!” (‘G‘afur G‘ulom «Vaqt».)

Bunda “abadiy” so‘zi “mustahkam, ulug‘vor” so‘zlari bilan sinonimlik hosil qiladi;

“Mangulikdan joy olgan - yashar abadiy,

Men-ku, Manguberdiman, olam

Ko‘klardanmi, suvdanmi, yo yer tagidan,

Balki sahro bag‘ridan, tog‘

Bir kun paydo bo‘laman shu yurt, shu yerda,

Yurt shaydosi ko‘milmas g‘urbatda - go‘rda.” (Maqsud Shayxzoda “Jaloliddin manguberdi”)

Bunda “abadiy” so‘zi “barhayot, o‘lmas” so‘zi bilan sinonimlik hosil qiladi;

“Ba‘zi xalqlarda dafn marosimida muzika chalishadi, bayroq ko‘tarib borishadi, ba‘zi xalqlarda osmonga o‘q uzishadi, boshqa xalqlarda esa gulchambarlar qo‘yib, marhumni gulga Edigey Bo‘ron bo‘lsa, ertaga tongdayoq popukli yopqich tashlangan Qoranorda Kazangapni so‘nggi abadiy manziliga — Ona Bayit mozoriga olib boradi...” (Chingiz Aytmatov “Asrga Tatigulik Kun”)

Bunda “abadiy” so‘zi “mozor” so‘zi bilan sinonimlik hosil qiladi;

“Lekin sanamdan ajralishga toqati yetmaydi. Chunki inson abadiy rohatda bo‘ladigan jannat ma‘shuqasiz “g‘amgoh”dir. Buni shoir shunday “Jannatdakim, el onda bo‘lur zindai jovid, Sensiz oni g‘amgoh degumdur dag‘i”. Oshiqning muhabbati shu qadar teran va azaliyki, u yor ko‘yida o‘lib ketsa ham uning ruhi qiyomatga qadar ma‘shuqaga hamroh. (Atoyi “Sensiz bu jahon ayshi” XV)

Bunda “abadiy” so‘zi “tinchlik, farovonlik” so‘zi bilan sinonimlik hosil qiladi;

“Agarda bular ichkilik ichmasa edi, dunyo va oxiratda aslo azob va mehnatda qolmas edi. Oh, haqiqatda, boyni o‘ldirgan va bu yigitlarni abadiy azobga giriftor qilgan beilmlikdur. Bizlarni xonavayron, bachagiryon va bevatan va bandi qilgan tarbiyasizlik va jaholatdur: beva- tanlik, darbadarlik, asorat, faqr-u zarurat va xorliklar hammasi ilmsizlik va betarbiyalikning mevasi va natijasidir.” (Mahmudxo‘ja Behbudiy “Padarkush”, 1911)

Bunda “abadiy” so‘zi “qiyin, mashaqqatli, tugamas” so‘zi bilan sinonimlik hosil qiladi;
 “Ne tong edi - izg‘irinlar zahrini totding,

Dasturxonda muztar qoldi qandu naboting.

Shahidlaring nomlaridir har bir raboting,

Abadiy yod erur asli adabiyoting,

Cho‘lpon, Fitrat, Qodiriy ham Usmonim, Vatan.” (Sirojiddin Sayyid “Vatan”)

Bu kontekstda “abadiy” so‘zi “unutilmas” so‘zi bilan sinonimlik hosil qiladi;

O‘lmoq, vafot etmoq, o‘chmas, ketmas, qoladigan, davomli, qiyin, barhayot, o‘lmas, adolatli, haqiqiy, mashaqqatli, tugamas, mustahkam, ulug‘vor, asl, chin, mozor, tinchlik, farovonlik, unutilmas kabi so‘zlar bir qaraganda “abadiy” so‘ziga sinonim bo‘la olmaydigandek, ammo konteks jihatidan qaraganda bu sinonim so‘zlar bo‘la oladi. Bu kontekstologik sinonimga birgina misol edi, ammo tilimizda bunga o‘xshash misollardan talayginasini uchratishimiz mumkin. Boshqa tillarda mavjud bo‘lgani kabi o‘zbek tilida ham bu kabi o‘zbek tili o‘rganuvchilari va bu tildan foydalanuvchilar uchun birdek zarur va foydali bo‘lgan kontekstologik sinonimlarning yirik bazasini yaratish bugungi kun tilshunos olimlari zimmasida turgan eng muhim masalalardan biridir.

Kompyuter texnologiyalari tobora rivojlanib borayotgan davrda sinonimlarning bazasini yaratishda yordam beradigan bir qancha sayt va dasturlar mavjud bo‘lib ulardan biri https://help.sap.com/docs/SAP_HANA_PLATFORM. Ushbu saytda hech qanday sinonim ta’rif kodini yozmasdan sinonimlarni yaratish va aniqlash uchun SAP Web IDE-dagi sinonim muharriridan foydalanish mumkin. Bunda avvaollo, asosiy ma’lumotlar bazasi foydalanuvchisi bilan foydalanuvchi tomonidan taqdim etilgan xizmatni aniqlab, loyihada talab qilinadigan barcha sxemalarga kirish uchun foydalanuvchi hisob ma’lumotlariga ega bo‘lishi kerak. Siz foydalanuvchi tomonidan taqdim etilgan xizmatni aniqlash uchun quyidagi sintaksisdan foydalanasiz.

1. Brauzerda SAP Web IDE-ni ishga tushirish.

2. Ma’lumotlar bazasi sinonimini yaratmoqchi bo‘lgan dastur loyihasini ko‘rsatish.

SAP Web IDE loyiha kontekstida dastur yaratadi . Agar sizda hali loyiha bo‘lmasa, uni yaratishning bir qancha usullari mavjud, masalan: uni import qilish, klonlash yoki noldan yangisini yaratish.

SAP Web IDE da Shablondan Yangi loyiha faylini tanlang.

Loyiha shablonining turini tanlang.

Hozirda faqat bitta turdagi loyiha shablonlari mavjud, xususan: Ko‘p maqsadli dastur loyihasi . Ko‘p maqsadli dastur loyihasini tanlang va Keyingi ni tanlang .

a. Yangi MTA loyihasi nomini kiriting (masalan, myAppva tasdiqlash uchun Keyingi ni tanlang).

b. Yangi MTA loyihasi tafsilotlarini ko‘rsating va tasdiqlash uchun "Keyingi" ni tanlang .

c. Yangi MTA loyihasini yarating; Finish ni tanlang .

3. Sinonimlarni yaratmoqchi bo‘lgan HDB modulini tanlang.

4. src jildiga o‘ting .

5. .hdbsynonym faylini yarating .

d. src jildining kontekst menyusida Yangi faylni tanlang .

e. Sinonimik deklaratsiya fayli nomini kiriting.

(Eslatma: Fayl nomiga qo‘shimcha sifatida .hdbsynonym qo‘shishingiz kerak.)

f. OK ni tanlang .

Asbob yangi sinonimlar muharririni ochadi, unda siz barcha kerakli sinonimlarni e'lon qilishingiz mumkin.

g. + ni tanlang .

h. Ma'lumotlar manbalarini topish dialog oynasida sinonimini e'lon qilmoqchi bo'lgan kerakli ma'lumot manbasini qidiring.

Bitta faylda bir nechta sinonimlarni e'lon qilishingiz mumkin.

(Eslatma: Shuningdek, ob'ekt nomi matn maydoniga ma'lumotlar manbai nomini to'g'ridan-to'g'ri kiritishingiz mumkin.)

i. Saqlash-ni tanlang .

6. HDB modulini yarating.

Qurilish jarayoni ma'lumotlar bazasi katalogida mos keladigan haqiqiy ob'ektlarni yaratish uchun dizayn vaqtidagi ma'lumotlar bazasi artefaktlaridan foydalanadi.

j. Modul kontekst menyusidan Build ni tanlang

Keying bosqichda sinonim yaratiladi

Sinonim nomi. Ushbu ob'ekt uchun ishlatadigan yangi nomni kiriting.

Sinonim sxemasi. Ushbu ob'ekt uchun ishlatadigan yangi nomning sxemasini kiriting.

Server nomi. Ulanish uchun server namunasini kiriting.

Ma'lumotlar bazasi nomi. Ob'ektni o'z ichiga olgan ma'lumotlar bazasini yozing yoki tanlang.

Sxema. ...

Ob'ekt turi. ...

Ob'ekt nomi.

Bunda ma'lumotlar bazasidagi sinonimlar nima, degan savol paydo bo'lishi mumkin. Sinonim - ma'lumotlar bazasi ob'ekti bo'lib, u mahalliy yoki uzoq serverda mavjud bo'lishi mumkin bo'lgan asosiy ob'ekt deb ataladigan boshqa ma'lumotlar bazasi ob'ekti uchun muqobil nom beradi. "Ma'lumotlar bazasi" so'zi kelgan har qanday iborada uning o'rniga "SCHEMA" kalit so'zidan sinonim sifatida foydalanish mumkin. To'g'ridan-to'g'ri ma'noda SCHEMA ma'lumotlar bazasining tuzilishini anglatadi.

Kontekstual sinonimlar bazasini yaratishda o'zbek tilidagi ma'lumotlar bazasi uchun yetarli bo'ladigan kerakli manbalarni topishda qiyinchiliklarga duch kelishimiz mumkin. Ba'zi sinonimlar har doim bir xil ma'noni anglatmaganligi tufayli ularning kontekstdagi farqlari aniqlanishi kerak, bu esa tahlilni murakkablashtiradi, bir sinonim turli kontekstlarda turli xil ishlatilishi mumkin, bu ularning o'zaro aloqalarini aniqlashda muammo tug'diradi. Til dinamikasi tufayli sinonimlar va ularning ishlatilishi vaqti-vaqti bilan o'zgaradi, bu esa bazaning yangilanishini talab qiladi, tashqari sun'iy intellekt va NLP vositalari yordamida sinonimlarni aniqlashda ba'zi cheklovlar bo'lishi mumkin, masalan, sifatli ma'lumotlarni aniqlashdagi qiyinchiliklar, uslubiy va grammatik xatolar. Bundan tashqari sinonimlar o'z uslubiga ko'ra farqlanadi (masalan, rasmiy, norasmiy), bu esa ularni to'g'ri tasniflashni talab qiladi, qolaversa, bazaning samaradorligini aniqlash uchun sinovlar o'tkazish va natijalarni baholashda ham qiyinchiliklar yuzaga kelishi mumkin. Bunda, ayniqsa, foydalanuvchilar bazadan qanday foydalanishini tushunish va ularni qoniqtiradigan interfeys yaratish ham juda muhim hisoblanadi.

Xulosa o'rnida shuni aytish mumkinki, sifat leksemalarining ma'no va ifoda jihatidan xilma-xilligi, ularning kontekstga bog'liq ravishda qanday o'zgarishi, tilshunoslikda muhim ahamiyatga ega. Shuningdek, sifat leksemalaridagi kontekstologik sinonimlar bazasining o'ziga xosligi va ahamiyati, jumladan, adabiyot va tilshunoslikdagi qo'llanilishi, til o'rganish jarayonida o'rganiladigan material sifatida alohida e'tiborga molikdir. Natijada, sifat leksemalarining kontekstual sinonimlari o'zbek tilining boyligini, uning semantik imkoniyatlarini kengaytiradi va tilni yanada rivojlantirishda muhim rol o'ynaydi. Maqola, shuningdek, kelgusidagi tadqiqotlar uchun yo'naltirishlar berib, o'zbek tilida sifat leksemalarining sinonimlar bazasini yanada chuqurroq o'rganish zarurligini ta'kidlaydi. O'z tilimizda ham rivojlangan tillarda mavjud

bo'lgani kabi kontekstologik sinonimlarning yirik bazasini yaratsak barcha o'zbek tili o'rganuvchilari va bu tildan foydalanuvchilar uchun birdek manfaatli bo'ladi degan umiddamiz.

Foydalanilgan adabiyotlar:

1. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, 11(19), 9117.
2. Abdurakhmonova, N., Alisher, I., & Sayfulleyeva, R. (2022, September). MorphUz: Morphological Analyzer for the Uzbek Language. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 61-66). IEEE.
3. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
4. Абдурахмонова, Н., & Абдувахобов, Г. (2021). О 'QUV LUG 'ATINI TUZISHNING NAZARIY METODOLOGIK ASOSLARI. МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИСКУССТВО СЛОВА, 4(6).
5. Abdurakhmonova, N., Shakirovich, I. A., & O'G'Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, 3(9), 146-156.
6. Ismailov, A. S., Shamsiyeva, G., & Abdurakhmonova, N. (2021). Statistical machine translation proposal for Uzbek to English. *Science and Education*, 2(12), 212-219.
7. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
8. Bateman, John and Cecile Paris. 1989. Phrasing a text in terms the user can understand. In Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89), volume 2, pages 1511–1517
9. Edmonds, Philip. 1999. Semantic Representations of Near-Synonyms for Automatic Lexical Choice. Ph.D. thesis, Computer Science Department, University of Toronto, Toronto.
10. Gove, Philip, editor. 1984. Webster's New Dictionary of Synonyms. Merriam-Webster. Hardt, Daniel and Owen Rambow. 2001.
11. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O 'ZBEK TILI ELEKTRON KORPUSIDA (<http://uzbekcorpus.uz/>) OG 'ZAKI MATNLAR KORPUSINI YARATISHNING NAZARIY VA AMALIY MASALALARI. *Academic research in educational sciences*, 3(3), 644-650.
12. Kubler S., Zinsmeister H. Corpus linguistics and linguistically annotated corpora. – New York: Bloomsbury, 2015. –P. 321.
13. Martin W. Developing Linguistic Corpora: A Guide to Good Practice, OxfordBooks. 2005. <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/>
14. Heike Z., Hinrichs E., Kübler S., Witt A. Linguistically annotated corpora: Quality assurance, reusability and sustainability / Corpus Linguistics: An International Handbook A. Lüdeling and M. Kytö (eds), Vol. 1, – Berlin: Mouton de Gruyter, –P. 759–76.

POLYSEMOUS NATURE OF HIGH-FREQUENCY PHRASAL VERBS

Otamurodova Salomat

Tashkent State Agrarian University
 Teacher at the department "Languages"
 E-mail: soloahmedovna@gmail.com

Annotation: The polysemous nature of English phrasal verbs presents a significant challenge in language comprehension and acquisition. Phrasal verbs, formed by combining verbs with particles (e.g., "take off," "break down"), often possess multiple, context-dependent meanings that differ greatly from their literal interpretations. This semantic variability arises from idiomatic usage, metaphorical extensions, and syntactic flexibility, making them difficult for both native speakers and language learners. Understanding their polysemy is crucial for grasping nuanced meaning in everyday communication and literature. This article highlights the linguistic complexity of phrasal verbs and emphasizes the need for contextual analysis in interpretation.

Keywords: polysemous, semantics, corpus tool, literal meaning, metaphorical, corpus data.

Annotatsiya: Ingliz tilidagi frazeologik fe'llarning polisemantik tabiati tilni tushunish va o'zlashtirish jarayonida jiddiy muammolarni keltirib chiqaradi. Fe'l va partikulalar (masalan, "take off", "break down") birikmasidan hosil bo'lgan frazeologik fe'llar ko'pincha kontekstga bog'liq bo'lgan, ko'pincha to'g'ridan-to'g'ri ma'nosidan keskin farq qiluvchi ko'p ma'noli birliklardir. Ularning semantik o'zgaruvchanligi idiomatik qo'llanilish, metaforik kengayishlar va sintaktik moslashuvchanlik tufayli yuzaga keladi. Bu esa ularni nafaqat til o'rganuvchilar, balki ona tilida so'zlashuvchilar uchun ham murakkablashtiradi. Frazeologik fe'llarning polisemiyasini chuqur anglash kundalik muloqot va adabiyotdagi nozik ma'nolarni tushunishda muhim ahamiyatga ega. Mazkur maqola frazeologik fe'llarning lingvistik murakkabligini yoritadi hamda ularni talqin qilishda kontekstual tahlil zarurligini ta'kidlaydi.

Kalit so'zlar: ko'p ma'noli, semantika/ma'no, korpus vosita, asl ma'no, metaforik/majoziy, korpus ma'lumot.

Аннотация: Многозначность английских фразовых глаголов представляет собой серьёзную проблему для понимания и освоения языка. Фразовые глаголы, образованные путем сочетания глаголов с частицами (например, "take off", "break down"), часто имеют несколько зависящих от контекста значений, которые значительно отличаются от их буквального смысла. Такая семантическая изменчивость обусловлена идиоматическим употреблением, метафорическим расширением значений и синтаксической гибкостью, что затрудняет их понимание как для носителей языка, так и для изучающих английский. Понимание их многозначности имеет важное значение для улавливания тонких смыслов в повседневной речи и литературе. В данной статье подчеркивается лингвистическая сложность фразовых глаголов и необходимость анализа контекста при их интерпретации.

Ключевые слова: многозначный, семантика, корпусный инструмент, буквальное значение, метафорический, корпусные данные.

Introduction

English phrasal verbs (PVs) present a unique challenge for learners and linguists due to their idiomatic nature, semantic opacity, and syntactic flexibility. A phrasal verb consists of a verb combined with one or more particles (prepositions or adverbs) that together form a meaning distinct from the individual words (e.g., "give up", "run into", "take off"). Traditional approaches to phrasal verbs have often focused on categorization or translation equivalents; however, corpus linguistics offers a dynamic methodology for uncovering patterns of usage and meaning in real-

world texts. By examining large, annotated corpora such as the British National Corpus (BNC), Corpus of Contemporary American English (COCA), or the more learner-focused Cambridge Learner Corpus (CLC), researchers can analyze how phrasal verbs are used across genres, registers, and contexts.

This approach reveals not only frequencies and collocations but also semantic nuances and idiomatic patterns. This paper explores how the semantics of English phrasal verbs are revealed through corpus data, focusing on usage frequency, polysemy, metaphorical extension, collocational patterns, and contextual disambiguation. Corpus analysis begins with frequency data, which reveals the most commonly used phrasal verbs. According to studies using COCA and BNC, a relatively small subset of phrasal verbs accounts for the majority of occurrences in spoken and written English. Gardner and Davies (2007), for example, identified 100 high-frequency phrasal verbs that make up 45% of all phrasal verb occurrences in COCA.

Such findings suggest a core set of semantically flexible verbs (e.g., "get", "take", "make", "go") tend to form phrasal verbs with various particles ("get up", "get over", "get by"). Frequency analysis not only indicates which PVs are important for learners to acquire but also gives clues to semantic productivity. High-frequency PVs are often polysemous, exhibiting multiple meanings depending on context. Here are some examples to some phrasal:

1. Take off

a. Literal meaning (removal):

He took off his shoes before entering the house.

Here, "take off" means to remove something, such as clothing.

b. Flight (aircraft):

The plane took off at exactly 9 AM.

In this context, "take off" refers to an aircraft becoming airborne.

c. Success or rapid growth:

Her career really took off after she moved to New York.

This figurative use of "take off" means to suddenly become successful or popular.

d. Leaving quickly or suddenly:

He saw the police coming and took off in the opposite direction.

Here, it means to leave a place suddenly or hurriedly.

The polysemous nature of high-frequency PVs adds richness and flexibility to English but also poses challenges. Without context, learners might misinterpret the meaning, especially when literal and idiomatic uses overlap. For example:

He turned up the hill vs. He turned up late

One describes physical movement, the other describes arrival.

Language learners benefit from:

Learning PVs in **contextualized examples**

Practicing them in **realistic dialogues**

Grouping meanings by **theme or usage** (e.g., emotional, directional, technical)

One of the most prominent semantic features of phrasal verbs is **polysemy**—the coexistence of multiple related meanings. A corpus allows us to track how a single phrasal verb, such as "take off", operates across various contexts:

- Literal: "The plane took off at noon."
- Figurative: "Sales took off after the advertisement aired."
- Causal: "He took off his shoes."

Corpus tools like concordancers and word sketch engines can generate KWIC (Key Word in Context) lines, which reveal these uses side by side, allowing researchers to observe how meaning shifts according to the co-text. This polysemy often follows metaphorical or metonymic principles. For example, "take off" originally described physical upward motion, which metaphorically extends to describe success or popularity. Cognitive linguistics provides a helpful framework for understanding phrasal verb semantics via metaphorical mapping. Corpus data

supports the idea that certain spatial particles (e.g., "up", "down", "out") contribute metaphorical meaning systematically. For example

- "Up" often indicates completion or increase ("eat up", "speed up").
- "Down" may suggest reduction or control ("slow down", "calm down").
- "Out" can signal emergence or exhaustion ("find out", "burn out").

By tracing how these patterns appear across thousands of instances, corpora reveal that particles are not random but semantically motivated. Moreover, corpus data shows that metaphorical extensions are not equally distributed—some PVs become idiomatic or fixed expressions, losing their compositional transparency. Another advantage of corpus analysis is its ability to reveal typical **collocational behavior**—which nouns, subjects, or grammatical structures tend to co-occur with certain phrasal verbs. This collocational data helps to disambiguate meanings and understand usage constraints. For example:

- "Break down" with machines: "The car broke down on the highway."
- "Break down" with emotions: "She broke down in tears."
- "Break down" with arguments: "Let me break down the problem."

Through corpus tagging and parsing, researchers can identify grammatical constructions (e.g., transitive vs. intransitive) and idiomatic clusters. These patterns help learners understand when and how PVs are used and which combinations are most natural. Moreover, collocational profiles differ across registers. For example, "pick up" in conversation might relate to physical actions ("pick up the phone"), while in academic writing, it might pertain to acquiring knowledge ("pick up on a theory"). Corpora like COCA allow register-specific filtering to highlight these nuances. A persistent challenge in phrasal verb analysis is **semantic disambiguation**—determining which sense of a PV is used in a given context. Corpus data supports this process by providing authentic usage examples that show how context clarifies meaning. Consider "put off":

- Delay: "He put off the meeting until Monday."
- Dissuade: "Her attitude really put me off."

Lexical and syntactic context often provide cues for disambiguation. In the first example, temporal expressions (e.g., "until Monday") suggest a postponement meaning. In the second, an animate object and an affective verb context (e.g., "attitude") point to aversion. Corpora annotated with part-of-speech tags, semantic roles, or word senses (e.g., SemCor) can be used to train machine learning models or assist in automatic sense disambiguation, which further supports semantic analysis and applications in NLP. Corpora can also be used diachronically to track **semantic changes** in phrasal verbs over time. For example, historical corpora like the Corpus of Historical American English (COHA) show how older senses of PVs faded or evolved and how new metaphorical uses emerged. This is particularly relevant for understanding the ongoing productivity of certain PVs in modern usage.

Cross-linguistic corpus studies also highlight how phrasal verbs pose difficulties for L2 learners whose native languages do not use similar verb-particle constructions. Learner corpora such as the International Corpus of Learner English (ICLE) reveal common errors and overgeneralizations, offering insight into how semantic opacity impacts acquisition.

Conclusion

Corpus analysis offers a rich, empirical approach to studying the semantics of English phrasal verbs. By examining frequency, polysemy, metaphorical extension, collocational patterns, and context-driven disambiguation, researchers can better understand how PVs function in actual usage. These findings are not only of theoretical linguistic interest but also have practical implications for lexicography, language teaching, and natural language processing. As corpora become more annotated and tools more sophisticated, future research can delve even deeper into the interface between form, meaning, and use in phrasal verbs—shedding further light on one of the most intricate areas of English semantics. High-frequency phrasal verbs like *take off*, *turn up*, *break down*, and *put up* exemplify how polysemy enriches the English language. Their diverse meanings across contexts make them powerful tools for expression, but also key focus areas in

English learning. Understanding the multiple meanings of PVs and recognizing contextual clues can significantly improve comprehension and fluency.

References:

1. Graham Workman "Phrasal Verbs: A Real English Book"
2. John Flower "Phrasal Verb Organiser"
3. Michael McCarthy., Felicity O'Dell. English Phrasal Verbs in Use"
4. "Pointing Out Frequent Phrasal Verbs: A Corpus-Based Analysis," published in *TESOL Quarterly*, Volume 41. Gardner., Davies's 200
5. Oxford University Press "Oxford Phrasal Verbs Dictionary for Learners of English"
6. Richard A. Spears. "Phrasal Verbs and Idioms"

O'ZBEK TILIDAGI ZAMONAVIY TARJIMA TIZIMLARINI YARATISHNING DOLZARB MASALALARI

Yo'ldoshboyeva Gulsanam Diyorbek qizi

Mirzo Ulug'bek nomidagi O'zbekiston Milliy Universiteti

Jurnalistika va o'zbek filologiyasi fakulteti.

Filologiya va tillarni o'qitish: o'zbek tili yo'nalishi

E-mail: gulsanamyoldoshboyeva@gmail.com

Annotatsiya: Maqolada hozirgi kunda mashina tarjimalarining ahamiyati, shu bilan birgalikda o'zbek tilidagi zamonaviy tarjima tizimlarini yaratishning dolzarb masalalari va bir nechta muammolarning tahlillari haqida so'z boradi. Jumladan, sinonim, omonim, ko'p ma'noli so'zlar tarjimasidagi qiyinchiliklar, ingliz va o'zbek tillaridagi ilmiy va rasmiy matnlar tarjimasidagi kamchiliklar, morfologik va sintaktik muammolar va tarjimada kontekstual yondashuv kabilar tilga olinadi. Mashina tarjimasi uchun zarur bo'lgan gap modellari haqida ham to'xtalangan bo'lib, bu borada qilingan tadqiqot ishlari tahlilga tortilgan. Shuningdek, "Google Translate" tarjima dasturi haqida ham to'xtalib o'tilgan bo'lib, uning yutuq va kamchiliklari ko'rsatib berilgan. Tadqiqotimizning asosiy maqsadi ilmiy izlanishlar yo'lida keng imkoniyatlar yaratish uchun terminologik parallel korpuslar asosida ilmiy va rasmiy matnlarni (o'zbek va ingliz tilida) tarjima qiladigan mashina tarjimasi texnologiyasining lingvistik asoslarini tahlil qilish, o'zbek tilidagi zamonaviy tarjima tizimlarini yaratishdagi muammolarni hal qilishdan iboratdir.

Kalit so'zlar: Mashina tarjimasi, Parallel korpus, gap modellari, NLP, Morfoanalizator, Sintaktik analizator, Tarjima lug'atlar.

Annotation: This article examines the significance of machine translation in the modern era and the pressing issues related to developing contemporary translation systems for the Uzbek language. In particular, it addresses challenges in translating synonyms, homonyms, and polysemous words, as well as difficulties in translating scientific and official texts between English and Uzbek. Additionally, the study explores morphological and syntactic issues and the importance of a contextual approach in translation. The article also discusses sentence models essential for machine translation and analyzes existing research in this field. Furthermore, the capabilities and limitations of the "Google Translate" system are evaluated. The primary objective of this research is to investigate the linguistic foundations of machine translation technology based on terminological parallel corpora for translating scientific and official texts (in Uzbek and English) and to resolve issues related to developing modern translation systems for the Uzbek language.

Keywords: Machine Translation, Parallel Corpus, Speech Models, NLP, Morphoanalyzer, Syntactic Analyzer, Translation Dictionaries.

Аннотация: В данной статье рассматривается значение машинного перевода в современном мире, а также актуальные вопросы и проблемы, связанные с созданием современных систем перевода для узбекского языка. В частности, анализируются трудности перевода синонимов, омонимов и многозначных слов, а также недостатки при переводе научных и официальных текстов между английским и узбекским языками. Кроме того, обсуждаются морфологические и синтаксические проблемы, а также необходимость контекстуального подхода в переводе. В статье рассматриваются модели предложений, необходимые для машинного перевода, и анализируются проведенные в этой области исследования. Также уделяется внимание системе «Google Translate», рассматриваются её преимущества и недостатки. Основная цель данного исследования — анализ лингвистических основ технологии машинного перевода, основанного на терминологических параллельных корпусах, для перевода научных и официальных текстов (на узбекском и английском языках), а также решение проблем, связанных с созданием современных систем перевода для узбекского языка.

Ключевые слов: Машинный перевод, параллельный корпус, модели речи, НЛП, морфоанализатор, синтаксический анализатор, переводческие словари

Insonlar 1940-yillardan boshlab mashina tarjimasini texnologiyasi bilan shug'ullanadilar va har bir yangi texnologik tadqiqot vaqt o'tishi bilan jarayonlarni asta-sekin yaxshilaydi. Mashina tarjimasini, sodda qilib aytganda, matnni bir tildan boshqa tilga tarjima qilish uchun sun'iy intellektdan foydalanishdir. Mavjud tarjimalarning katta ma'lumotlar bazasidan olingan algoritmlar, andozalar va til modellaridan foydalangan holda, mashina tarjimasini texnologiyasi hech qanday inson ishtirokisiz tezda ko'p sonli tarjima qilingan matnlarni yaratish mumkin.

Ilm-fan rivojiga sabab bo'layotgan va bilim sohasi infratuzilmasiga ijobiy ta'sir ko'rsatadigan, asosan, butun dunyo bo'ylab yuqori tarmoqli kenglikka ega bo'lgan axborot uzatish texnologiyalari orqali bog'langan axborot texnologiyalarining yaqqol yutug'idir. Keng ma'noda aytganda, mamlakatimizda mustaqillikka erishgach, jamiyatning har bir tarmog'i tizimi isloh qilindi. Qolaversa, ilg'or o'zgarishlarga aylangan ta'lim tizimi o'sha yillardan boshlab bir vaqtning o'zida taraqqiy etib bordi. Xususan, hisoblash texnologiyalarining keng miqyosda tarqalishi, xususan, internet tizimidan foydalanish imkoniyatlarining ochilishi ilm-fandagi jiddiy muammolarga duch kelish uchun dunyo eshigini ochdi. Kompyuter bilan kesishgan fanlarning aksariyat yo'nalishlari paydo bo'ldi va davom etdi. Mashina tarjimasini nazariyasi kibernetika, tilshunos, faylasuf, matematik, dasturchi va psixologlar kabi qator soha vakillarining diqqat markazida bo'lgan.

Bugunga qadar mashina tarjimasini sohasida bir qancha olimlar tomonidan ilmiy izlanishlar olib borilgan. N. Abdurahmonova "Mashina tarjimasining lingvistik ta'limoti" nomli monografiyasi, A.But, R. Richans, J. Xatchins, P. Braun (AQSh); G.G.Belonogov, J.Allen, Z.Shalyapina, N.D. Andreev, I.A. Melchuk, V.Yu. Rozensveyg, Yu.N. Marchuk, R.G. Piotrovskiy, Yu.A. Motorin, K.B. Bektaev, A.N. Belyaev, I.K. Belskiy, A.V. Zubov, G.E. Miram, L.L. Nelyubin, V.I. Perebinyos, V.A. Chijakovskiy, Ye.A. Shingarev, R.G. Kotov, Babushkina N.V, O.Yu. Mansurova, A.S. Panina, A.A. Xoroshilov (Rossiya); M. Nagao (Yaponiya); A.Vaxer (Estoniya); Fransiyada J.Astrouni; R.Sinha, A.Jain (Hindiston); B.Bleyzer, U.Shvol, A.Storrer (Germaniya) singari tadqiqotchilarning ishlarini sana o'tadi⁶³. Massachusetts Texnologiya Institutining olimi Philipp Koehn "Pharaoh" ("Fir'avn") mashina tarjimasini mexanizmi orqali frazalarga asoslangan statistik mashina tarjimasini modellarini tavsiflab beradi⁶⁴.

Mashina tarjimasini sohasida birinchi ilmiy-amaliy konferensiya 1952-yilda Massachutes texnika universitetida o'tkaziladi va 1954-yilda NyuYorkdagi Jorjtaun universiteti bilan IBM kompaniyasi hamkorligida birinchi IBM II rus tilidan ingliz tiliga tarjima qilishga mo'ljallangan mashina tarjimasini dasturi yaratiladi. Bu dastur L.Dorster rahbarligida o'tkazilib, kimyo sohasi doirasida 250 ta leksik birlik hamda 6 ta grammatik qoidalar bilan chegaralgan edi. ("Dorster tajribasi" yoki "Jorjtaun sinovi").

O'zbek tili uchun mashina tarjimasiga doir deyarli fundamental izlanish olib borilmagan edi. N. Abduraxmonova "Mashina tarjimasining lingvistik ta'limoti" monografiyasida M.Xakimov ishlarida kengaytirilgan matematik modelga asoslanilgan mashina tarjimasini texnologiyasi bo'yicha tadqiqotlar amalga oshirilganligi, biroq o'zbek tilidagi ma'lumotlarning yetarli emasligi yaratiladigan dasturiy ta'minotlarga ta'sir ko'rsatganligi qayd etib o'tilgan⁶⁵.

Abduraxmonova tomonidan "Inglizcha matnlarni o'zbek tiliga tarjima qilish dasturining lingvistik ta'minoti (sodda gaplar misolida)" mavzusidagi falsafa doktori (PhD) ilmiy darajasini

⁶³ Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) ildis. aftoref." (2018).

⁶⁴ Philipp Koehn. A Beam Search Decoder for Phrase-Based Statistical Machine Translation Models. Machine Translation: From Real Users to Research 6th Conference of the Association for Machine Translation in the Americas, AMTA 2004. Washington, DC, USA, September/October 2004

⁶⁵ Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) ildis. aftoref." (2018). // Мухамедов С.А., Пиотровский Г.Г. Инженерная лингвистика и опыт системно – статистического исследования узбекских текстов. –Т.: Фан, 1986; Махмудов М.А., Пиотровская А.А., Садыков Т. Система машинного анализа и синтеза тюркской словоформы // Переработка текста методами инженерной лингвистики. –Минск, 1982.

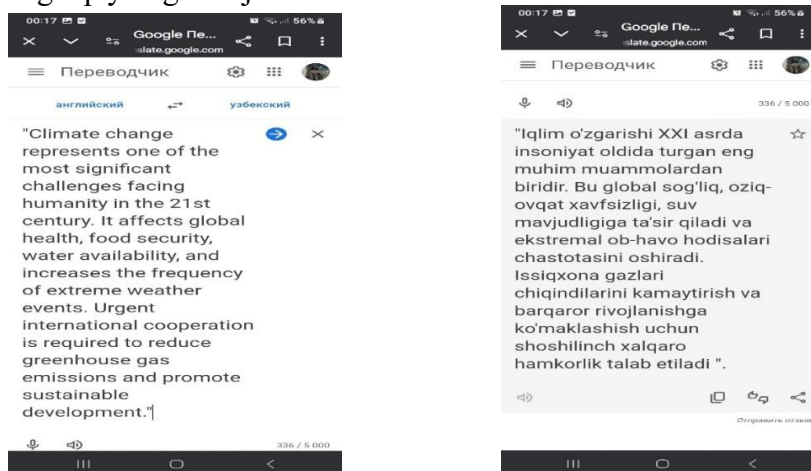
lish uchun yozilgan (2018) dissertatsiyasida o'zbek tilining lingvistik modellari, sintaktik strukturalar orqali tarjima qilish, qolaversa, transfer usuli qardosh bo'lmagan tillar uchun mos ekanligi ilmiy jihatdan o'rganilgan⁶⁶.

Mashina tarjimasi texnologiyasi rivojlanib borayotgan bo'lsa-da, sinonim, omonim va ko'p ma'noli so'zlarni tarjima qilishda hali ham qiyinchiliklarga duch keladi. Mashina tarjimasi sinonimlarni kontekstga qarab ajrata olmaydi va noto'g'ri sinonimni tanlashi mumkin. Misol uchun "ulkan" va "bahaybat" sinonimlarini olsak. "Ulkan yutuq" deyish mumkin, lekin "bahaybat yutuq" biroz noaniq eshitiladi. "Tez" va "shoshilinch" sinonimlarida esa "tez yordam" va "shoshilinch yordam" iboralarini kontekstga qarab farq qiladi. Bunday noaniqliklar kelib chiqmasligi uchun mashina tarjimasiga kontekstual sinonimlarni qo'shish lozim. Tarjima tizimi sinonimlarning eng mos keladigan variantini tanlay olish uchun leksik analizdan foydalanishi kerak. O'zbek tilidagi sinonimlarning kontekstga bog'liqligini aniqlash uchun katta hajmdagi matnlarni o'rganuvchi tizimlar kerak va ularni ishlatilishini chuqur o'rgatish uchun transformer modellaridan foydalanish lozim.

Omonimlar bir xil yoziladigan, lekin turli ma'nolarga ega bo'lgan so'zlardir. Mashina tarjimasi bu so'zlarning to'g'ri kontekstini aniqlashda xatolarga yo'l qo'yishi mumkin. Masalan, "ochiq" so'zi (yopilmagan) va (ochiq gapirish) kabi turli ma'noga ega. Tarjimada xato kelib chiqmasligi uchun mashina tarjimasi butun jumlani yoki paragrafni analiz qilishi kerak, faqat so'zni emas. Shu bilan birga semantik tarmoqlardan foydalanish ya'ni so'zlarning turli kontekstlaridagi ma'nolarini aniqlash uchun WordNet kabi ma'lumot bazalaridan foydalanish lozim. Transformer modellarini qo'llash ham tarjimani aniqroq tarjima qilish uchun qo'l keladi.

Ko'p ma'noli so'zlar tarjimasida ham bir qancha muammolar vujudga keladi. Masalan, "bosh" – tana a'zo va yetakchi, bosh direktor ma'nolarini anglatishi mumkin. Vujudga kelayotgan muammolarni bartaraf etish uchun qo'shimcha semantic tahlil qilish ya'ni NLP (Natural Language Processing) yordamida ko'p ma'noli so'zlarning ishlatilish ehtimolini hisoblash kerak. Shu bilan birga transformer modellaridan foydalangan holda matn ichidagi ko'o ma'noli so'zlarning aniq tarjimasini tanlash tizimini yaratish zarur.

Mashina tarjimasida ko'plab tadqiqotlar olib borilganiga qaramay o'zbek tilining avtomatik tarjima tizimi mavjud emas. Bir tildagi matnlarni o'zbek tiliga tarjima qilish uchun "Google Translation" dasturidan keng foydalaniladi. Bu dastur bir qancha qulayliklarga ega bo'lish bilan birga, tarjimada xatoliklar ham uchraydi. Bu, ayniqsa, ilmiy va rasmiy matnlar tarjimasida ko'p kuzatiladi. Misol tariqasida Morsy Cheikhrouhouning **"The impact of Climate Change on Earth and Human Society During the Next Century"** maqolasidan parcha tarjima qilib ko'rdik va bizga quyidagi natijani berdi:



1-rasm. Inglizcha ilmiy maqoladan olingan parcha; 2-rasm. Inglizcha ilmiy maqoladan olingan parchaning Google Translate yordamida qilingan o'zbekcha tarjimasi.

⁶⁶ Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) ildis. aforef." (2018).

Climate change represents one of the most significant challenges facing humanity in the 21st century. It affects **global health**, food security, **water availability** and increases the frequency of extreme weather events. Urgent international cooperation is required to reduce **greenhouse gas emissions** and **promote sustainable** development.

Iqlim o'zgarishi XXI asrda insoniyat oldida turgan eng muhim muammolardan biridir. Bu **global sog'liq**, oziq-ovqat xavfsizligi, **suv mavjudligiga** ta'sir qiladi va ekstremal ob-havo hodisalari chastotasini oshiradi. **Issiqxona gazlari chiqindilarini** kamaytirish va **barqaror rivojlanishga ko'maklashish** uchun shoshilinch xalqaro hamkorlik talab etiladi.

Tarjima natijasida ko'rinib turibdiki, ayrim so'zlar va terminlar matn ma'nosiga va gap mazmuniga to'g'ri kelmaydigan holda tarjima qiliyapti. Bu esa kontekstdagi ma'nosini anglashilmasligiga olib keladi. Buning asosiy sabablaridan biri "Google Translate" dasturida o'zbek tili ma'lumotlar bazasining to'liq emasligidadir. Shu bilan birga terminologik noaniqliklar yuzaga keladi. Ilmiy va texnik terminlar ba'zan noto'g'ri yoki umumiy ma'noda tarjima qilinadi. Masalan, "resistance" fizikada "qarshilik" bo'lishi kerak, lekin ba'zida "qarshilik ko'rsatish" deb tarjima qilinadi. Sababi kontekstni tahlil qilish mexanizmi yetarlicha chuqur emas. Grammatik strukturada esa o'zbek tilidagi gap qurilishi ingliz tiliniki bilan farqlanadi. "Google Translate" ko'pincha inglizcha struktura asosida o'zbekcha gaplar tuzadi, bu esa g'ayritabiiy ko'rinadi. Masalan, "He conducted a research..." – "U tadqiqot o'tkazdi" o'rniga "U bir tadqiqotni bajardi" deb tarjima qilinishi mumkin. Bunday kamchiliklar kelib chiqmasligi uchun maxsus sohalar uchun tarjima modellarini yaratish, masalan, huquqiy, tibbiyot yoki texnik sohalarga moslashtirilgan tarjima tizimlarini yaratish lozim. Kontekstni aniqlash algoritmlarini takomillashtirish – WSD (Word Sense Disambiguation) va Attention mexanizmlarini kuchaytirish ham zarur.

Philipp Koehn tavsiflab bergan statistik mashina tarjimasi modellarini va "Google Translate" ni analiz qilgan holda O'zbek tilshunosligida korpusga asoslangan mashina tarjimasi texnologiyasini joriy etish zarur. Buning uchun katta hajmdagi korpus kerak bo'ladi. Bunda Nilufar Abduraxmonova rahbarligida yaratilgan <https://uzbekcorpus.uz/> dan foydalanish maqsadga muvofiq.

Matnlarni terminologik, struktur, morfologik, sintaktik tomondan to'g'ri tarjima qiladigan avtomat tarjima dasturi texnologiyasini yaratish uchun katta hajmdagi korpus bilan bir qatorda quyidagilar kerak bo'ladi:

1. Parallel korpus

2. Morfo analizator

3. Sintaktik analizator

4. Tarjima lug'atlar

Tarjima dasturi to'g'ri ishlashi uchun matnlarni o'xshashlik jihatidan ehtimollilik modellarini yaratib chiqishimiz zarur bo'ladi. Modellashtirishning quyidagi turlari mavjud:

- So'z asosida modellashtirish, ya'ni so'z segmentlari bo'yicha muqobil variantlarni yaratish;

- So'z birikmalari yoki frazalar bo'yicha muqobillashtirish;

- Gaplarni o'zaro muqobillashtirish.

Avvalo, yaratilayotgan mashina tarjimasi uchun taxminan nechta gap modeli kerak bo'lishini aniqlab olish lozim. O'zbek tilshunosligida gap modellari bir qancha olimlar tomonidan o'rganilgan bo'lib, xususan, Nilufar Abduraxmonova aynan tarjima dastur uchun ingliz va o'zbek tillaridagi gap modellarini tadqiq etish maqsadida mashina tarjimasi nazariyasi bo'yicha ilmiy izlanishlar olib borgan va "Mashina tarjimasining linvistik ta'minoti" nomli monografiyasining to'rtinchi bobini aynan mashina tarjimasi tizimida gap modellarini yaratish nazariyasiga bag'ishlagan⁶⁷.

⁶⁷ Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) ildis. aforef." (2018).

Monografiyada 50 dan ortiq o'zbek-ingliz tillaridagi gap modellari keltirilgan bo'lib, ular ingliz tilidagi ot kesimli sodda gap va fe'l kesimli sodda gap modellari, shuningdek, inglizcha matnlardagi ega va kesim moslashuvining o'zbek tilidagi modellari hisoblanadi. Tadqiqot ishida kesimi o'timisiz fe'ldan iborat bo'lgan asosiy gap modellarining 10 ta, kesimi yordamchi fe'l bilan hosil bo'lgan asosiy gap modellarining 14 ta, kesimi o'timli fe'l bo'lgan asosiy gap modellarining 13ta, to be yordamchi fe'li ishtirok etadigan gap shakllarining 20 dan ortiq turi ko'rsatib berilgan. Mashina tarjimai to'g'ri ishlashi uchun qo'shma gap modellari ham kerak bo'ladi. Ingliz va o'zbek tillaridagi qo'shma gap modellarini avtomatik tarjima uchun ishlab chiqish oldimizda turgan vazifalardan biri bo'lib hisoblanadi.

O'zbek tilidagi zamonaviy tarjimai tizimlarini yaratish va undagi kamchilik va muammolarni bartaraf etish o'zbek olimlar, umuman, ilmiy tadqiqot ishini olib boruvchilar uchun keng imkoniyat yaratadi. Yuqoridagilarni hisobga olib, yutuq va kamchiliklarni tahlil qilgan holda o'z tadqiqot ishimizda ilmiy terminlar va ilmiy matnlar bazasini kengaytirib, ilmiy parallel matnlarni to'ldirish hisobiga boyitib, tarjima tizimidagi kamchilik va muammolarni hal etgan holda o'zbek tilshunosligida zamonaviy tarjima tizimlarini yaratishni maqsad qildik.

Foydalanilgan adabiyotlar:

1. Aripov, M., Sharipbay, A., Abdurakhmonova, N., Razakhova B.: Ontology of grammar rules as example of noun of Uzbek and Kazakh languages. In: Abstract of the VII International Conference "Modern Problems of Applied Mathematics and Information Technology -Al-Khorezmiiy 2018", pp. 37–38, Tashkent, Uzbekistan (2018)
2. Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) ildis. aftoref." (2018).
3. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
4. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
5. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. In Proceedings of the International Conference on Language Technologies for All (LT4All) 2019.
6. Philipp Koehn. A Beam Search Decoder for Phrase-Based Statistical Machine Translation Models. Machine Translation: From Real Users to Research 6th Conference of the Association for Machine Translation in the Americas, AMTA 2004. Washington, DC, USA, September/October 2004
7. Jalil, Masita & Ismailov, Alisher & Abd Rahim, Noor Hafhizah & Abdullah, Zailani. (2017). The Development of the Uzbek Stemming Algorithm. Advanced Science Letters. 23. 4171-4174. 10.1166/asl.2017.8332

ПРИМЕНЕНИЕ КОРПУСНОГО ПОДХОДА В ИССЛЕДОВАНИЯХ ЯЗЫКА

Хидирова Гулнора

кандидат филологических наук, доцент
Университет Маъмун

Аннотация: Статья посвящена описанию корпусного подхода и проведению на его основе различных исследований. В настоящее время характерно интенсивное изучение современной лингвистики посредством исследования корпусных данных разных языков. Корпуса различного рода стали неотъемлемым показателем объективности и достоверности полученных результатов. При корпусном методе обрабатывается значительный по объёму массив данных, вследствие чего действительно решается задача репрезентативности материала. В работе с корпусами возможны автоматический поиск информации и сплошной анализ материала. Представительность корпусных данных определили основной принцип корпусного метода. Данный метод подразумевает применение статистических способов и приемов обработки данных, поэтому корпусный подход означает применение количественного/квантитативного метода в исследовании проблем языка.

Ключевые слова: корпус, метод, исследование, язык, массив данных, репрезентативность, автоматический поиск информации, сплошной анализ, представительность, статистические способы, количественный, квантитативный метод.

Abstract: The article is devoted to the description of the corpus approach and the implementation of various studies based on it. At present, intensive study of modern linguistics is characteristic through the study of corpus data of different languages. Corpora of various kinds have become an integral indicator of the objectivity and reliability of the results obtained. The corpus method processes a significant amount of data, as a result of which the problem of representativeness of the material is effectively solved. When working with corpora, automatic information search and continuous analysis of the material are possible. The representativeness of corpus data determined the main principle of the corpus method. This method implies the use of statistical methods and techniques for data processing, therefore the corpus approach means the use of a quantitative method in the study of language problems.

Keywords: corpus, method, research, language, data array, representativeness, automatic information retrieval, continuous analysis, representativeness, statistical methods, quantitative, quantitative method.

Введение. На современном этапе в Узбекистане приняты значимые решения по дальнейшему сохранению и развитию узбекского языка, повышению его роли и статуса как национального и государственного. «Родной язык как символ национального самосознания и гордости, безусловно, служит для нас источником огромной силы и вдохновения», - отмечает Президент Ш.М. Мирзиёев [Из речи Президента, 2021].

В утвержденной в 2020 году Концепции развития узбекского языка определены такие приоритеты, как открытие новых направлений «Компьютерная лингвистика» и «Прикладная филология» в вузах филологического профиля, а также создание Национального электронного корпуса узбекского языка [Указ Президента Республики Узбекистан, 2020].

Разработки национальных корпусов во всем мире и прикладные изыскания в корпусной лингвистике убедительно доказывают, что корпус является не только инструментом и методом в исследовании языка, но также является способом сохранения языка и нации [N.Abduraxmonova, 2021].

Основная часть. Отметим, что начало XXI века знаменуется мощным прорывом информационных технологий в сфере филологических наук: стало возможным обучение языку посредством так называемых больших массивов текстов. Компьютерные технологии способствовали развитию практики создания корпусов, в связи с чем в 80-е годы оформляется самостоятельный раздел языкознания – корпусная лингвистика, которая и занимается разработкой, созданием и использованием текстовых корпусов. Созданные корпусы текстов вобрали в себя широкий спектр языковых данных, таких как словари, художественные произведения, научно-технические тексты, современные тексты определённой тематики и т.п.

Российский лингвист В. Плунгян, стоявший у истоков создания Национального корпуса русского языка, в своих рассуждениях о новом взгляде на язык, отмечал, что «если мы хотим назвать такую область лингвистики, которая по определению является суперсовременной, то первое, что приходит в голову, – это как раз лингвистика корпусов» [В. Плунгян, 2013].

Современные корпуса представляют собой большую электронную коллекцию текстов. Тексты упорядочены по определенным параметрам и служат для исследователей в качестве иллюстративного материала. Тексты подвергаются структурной и лингвистической разметке, посредством чего компоненты текста снабжаются определенными характеристиками (структурными и собственно лингвистическими). Размеченные таким образом тексты конвертируются в структуру корпусного менеджера (информационной системы управления), которая состоит из программных средств для многоцелевого поиска данных в корпусе и получения статистической информации.

Захаров К.Л. определял целесообразность создания и смысл использования корпусов следующими факторами: «1) достаточно большой (репрезентативный) объем корпуса гарантирует типичность данных и обеспечивает полноту представления всего спектра языковых явлений; 2) данные разного типа находятся в корпусе в своей естественной контекстной форме, что создает возможность их всестороннего и объективного изучения; 3) однажды созданный и подготовленный массив данных может использоваться многократно, многими исследователями и в различных целях» [К.Л. Захаров, 2005].

Результаты поиска представляются корпусным менеджером в виде так называемого конкорданса, представляющий собой полный список всех употреблений слова в контекстном окружении. Параллельно с созданием и развитием различных корпусов начинает активно формироваться и использоваться в прикладной лингвистике характерный метод корпусного анализа. Наша статья посвящена описанию данного корпусного подхода и проведения на его основе различных исследований.

Исследования языка могут проводиться в соответствии с аспектами и категориями, изучение которых состоит в приоритете. Это может быть, изучение проблемы одного уровня, как лексика, морфология, семантика, синтаксис, грамматика, стилистика или дискурс-анализ [К. М. Шилихина, 2014]. Исследуемое явление может рассматриваться также в сочетании этих уровней [Методы когнитивного анализа семантики слова, 2015].

Надо подчеркнуть, что «корпусное исследование – это больше, чем методика анализа» [В.А. Плунгян, 2009]. Созданные языковые корпуса отличаются большим массивом данных, корпуса по сути является коллекцией языковых элементов во всем своем жанровом и стилевом многообразии. Их представительность очевидна, вследствие чего, исследования, проводимые в корпусах, позволяют получать более точные результаты. Нужно отметить, что к настоящему времени лингвисты, проводящие исследования на основе корпусных методов, являются также разработчиками этих корпусов. Показательным примером тому может служить Национальный корпус русского языка (НКРЯ) [<https://ruscorpora.ru/>].

Рассмотрим возможности метода корпусных исследований языка. Для десятилетий XXI века характерно интенсивное изучение современной лингвистики посредством

исследования корпусных данных разных языков. Корпуса различного рода стали неотъемлемым показателем объективности и достоверности полученных результатов. Выше отмечалось, что при корпусном методе обрабатывается значительный по объёму массив данных, вследствие чего действительно решается задача репрезентативности материала. Также в работе с корпусами возможны автоматический поиск информации и сплошной анализ материала. Функциональный подход к семантике и сочетаемости языковых единиц является определяющим, характерным критерием для исследований на основе корпусов текстов.

Представительность корпусных данных и определили основной принцип корпусного метода. Данный метод подразумевает применение статистических способов и приемов обработки данных, поэтому корпусный подход означает применение количественного/квантитативного метода в исследовании проблем языка.

Корпусы данных позволяют получить различные статистические сведения относительно языковых и речевых единиц. Это могут быть справки о частоте словоформ, лексем, грамматических категорий и т.п. Корпусному менеджеру также доступно представление статистических справок другого формата, таких как изменение контекстов в различные периоды времени или проследить совместную встречаемость лексических единиц и др. Это позволяет исследовать динамику изменения лексического состава языка, анализировать лексико-грамматические характеристики авторского стиля, текстов разных жанров. В современных корпусных исследованиях на основе обширных фактических данных корпусов возможно построение и уточнение грамматик.

Наряду с этим, в настоящее время электронный корпус текстов успешно используется в лексикографии. Во многих странах национальные корпуса стали инструментом в создании различных диахронических и синхронических словарей. Примером могут служить словари, созданные на основе Национального корпуса русского языка [<http://dict.ruslang.ru>].

Следующим функциональным потенциалом корпуса является возможность его использования в учебном процессе. Убедительным примером такого учебного корпуса, на наш взгляд, является Корпус современного американского варианта английского языка Марка Дэвиса (Corpus of Contemporary American English или корпус COCA) [О.Ю. Павлова, 2021]. В частности, в корпусе COCA предложен ряд заданий по «проверке лексической сочетаемости, поиску синонимов и проверке их сочетаемости с другими языковыми единицами, изучению деривационных возможностей слов, составлению тематических списков ключевых слов, поиску иллюстративного материала и работе с конкордансом» [О.Ю. Павлова, 2021].

Корпусный метод также успешно применяется при обучении и в исследованиях перевода. В создаваемых параллельных подкорпусах большого национального корпуса собираются тексты (языка-оригинала) и их переводы. К примеру, в НКРЯ созданы и функционируют более 30 двуязычных корпусов (английский, болгарский, белорусский, испанский, китайский, корейский и др.) [<https://ruscorpora.ru/>]. Обучение переводу на основе корпусов позволяет овладеть иностранным языком на соответствующем профессиональном уровне.

При анализе перевода посредством корпусного подхода лингвистами решается проблема, связанная с выбором эквивалентного слова, сочетания или предложения.

Выводы. Таким образом, корпуса – это коллекции текстов, которые были собраны в соответствии с определенными критериями, и корпусные исследования предполагают их автоматическое изучение. Все перечисленные выше функции корпуса основаны на применении количественного (квантитативного, статистического) метода представления результатов анализа: метод когнитивного анализа семантики слова; метод сплошного анализа словоформ. Вследствие этого использование представительного объема корпусных

данных является определяющим фактором в применении корпусного подхода для лингвистических исследований.

Метод сплошного анализа несловарных словоформ с помощью корпусных инструментов позволяет выявить лексико-семантические и стилистические особенности текстов разных стилей. Корпусный подход, или метод лингвистического исследования, основанный на корпусах текстов, ориентирован на прикладное изучение языка, его функционирование в реальных средах и текстах, что важно для исследования и преподавания языка.

Использованная литература

1. Abduraxmonova N. Kompyuter lingvistikasi (darslik), Toshkent: Nodirabegim, 2021. -B.400.
2. Захаров К.Л. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005. – 48 с.; стр.3.
3. Из речи Президента, посвященной празднованию узбекского языка. 20 октября 2021 года.
4. Концепция развития узбекского языка и совершенствования языковой политики в 2020 — 2030 годах. Указ Президента Республики Узбекистан от 20 октября 2020 года № УП-6084.].
5. Методы когнитивного анализа семантики слова: компьютерно-корпусный подход / под общ. ред. В. И. Заботкиной. - М.: Языки славянской культуры, 2015. -344 с.
6. Национальный корпус русского языка. - Режим доступа: www.ruscorpora.ru
7. Павлова О.Ю. Использование языковых корпусов в обучении иностранному языку. <https://cyberleninka.ru/article/n/ispolzovanie-yazykovyh-korpusov-v-obuchenii-inostrannomu-yazyku>
8. Плунгян Владимир. Корпусная лингвистика. 14.01.2013. <https://postnauka.ru/video/7783>
9. Плунгян В. А. Почему современная лингвистика должна быть лингвистикой корпусов. Лекция, прочитанная в рамках проекта «Публичные лекции». - Режим доступа: <http://polit.ru/article/2009/10/23/corpus/>
10. Шилихина К. М. Использование корпусов в исследованиях дискурса / К. М. Шилихина // Вестник Воронеж. гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация. - 2014. - № 3. - С. 21-26.
11. <http://dict.ruslang.ru/> Гришина Е. А., Ляшевская О. Н. Грамматический словарь новых слов русского языка; Ляшевская О. Н., Шаров С. А. Новый частотный словарь русской лексики; Кустова Г. И. Словарь русской идиоматики. Сочетания слов со значением высокой степени; Бирюк О. Л., Гусев В. Ю., Калинина Е. Ю. Словарь глагольной сочетаемости непредметных имен русского языка

ZAMONAVIY TILSHUNOSLIK VA TARJIMASHUNOSLIKNING DOLZARB MUAMMOLARI

Jo`raboyeva Mohidil Javlon qizi

Toshkent Kimyo Xalqaro universiteti Namangan filiali,
“Ingliz tili” yo‘nalishi 3-kurs talabasi

Ilmiy rahbar: **Hazratqulov Oybek**

Annotatsiya: Ushbu maqola Oliy ta’limni raqamlashtirish va zamonaviy pedagogikalar bilan xorijiy tillarni o‘qitish va ta’lim tizimini raqamlashtirish jarayoni hamda ilm-fandagi innovatsiyalar, insoniyatga sog‘lom- qulay muhitni yaratish yo‘llari haqida yoritilgan.

Kalit so‘zlar: Oliy ta’limni raqamlashtirish, xorijiy tillarga bo‘lgan talablar, zamonaviy pedagogika va raqamli muhit.

Аннотация: В статье рассматриваются вопросы цифровизации высшего образования и процесса обучения иностранным языкам с использованием современных педагогических подходов и цифровизации системы образования, а также инновации в науке и пути создания здоровой и комфортной среды для человечества.

Ключевые слова: цифровизация высшего образования, требования к иностранным языкам, современная педагогика и цифровая среда.

Annotation: this article discusses the digitization of higher education and the process of teaching foreign languages with modern pedagogies and digitizing the education system, as well as innovations in science and ways to create a healthy and comfortable environment for humanity.

Keywords: digitalization of higher education, requirements for foreign languages, modern pedagogy and digital environment.

Kirish

Hozirgi kunda juda tez sur’atda rivojlanib kelayotgan 21-asrning dolzarb muammolaridan zamonaviy tilshunoslik va tarjimashunoslikning o‘zbek tilshunosligining kompyuter bilan bog‘liq sohalarini rivojlantirish kerak bo‘lgan sohalar; axborot matnlaridagi qoliplik, kompyuterning tarjima lug‘ati va uning izohli shaklini ham yaratish, o‘zbek tilining milliy korpusini ishlab chiqish, kompyuterda inglizcha-o‘zbekcha tarjima dasturlarini ishlab chiqish va kompyuterdagi matnlarni tahrir qilishni dasturlarini yaratishi va uning foydalanishdagi funksiyalarni rivojlantirish “kompyuter texnologiyalari asri”ning yana bir dolzarb, ilm-fan rivoji uchun qo‘yilgan yo‘nalishlaridan biri bu-axborot texnologiyalari ta’lim tizimidagi o‘quv jarayoniga tatbiq etish hamda o‘zbek tilshunosligiga olib kirilgan yangi kompyuter sohasi rivoji kompyuter lingvistikasi sohasini rivojlantirishdir. Keng miqyosdagi kompyuter lingvistikasi 1960-yilda AQSHda shakllangan. Eng muhim tendensiyalari fan, inson va texnologik mashinalari kabilarning o‘zaro aloqa munosabatidir. O‘zbekiston Respublikasining oliy ta’lim tizimini avlodlar o‘rtasida raqamli tizim bilan moslashtirish konsepsiyasida. “O‘zib kelayotgan yosh avlodni raqamli sanoatga o‘tishi-mamlakatni raqamli transformatsiyasini amalda qo‘llashga shuningdek, inson kapitalini tez sur’atda rivojlanishi”ga olib keladi. B.E.Starchenkoning ta’limotiga ko‘ra, ta’lim tizimini raqamlashtirish deganda an’anaviy ta’lim tizimidan yangi-Raqamli ta’limga o‘tish jarayoni tushiniladi. Ta’lim jarayoni kompyuterlashtirilgan ekan, bevosita kompyuter tizimidagi ma’lumotlar bazasi kompyuter lug‘atlari, metod va shakllarini qayta ko‘rib chiqish va yangi bilim va ko‘nikmalarni yanada boyitish. avtomatik tarjima, avtomatik tahrir, til o‘qitish dasturlarini qondirish kabi sohalar uchun kompyuter lingvistikasi sohasida izlanishlar olib borishni taqozo etadi.

Oliy ta'limni raqamlashtirishda zamonaviy mutaxassiszlarni jalb qilish va axborot-kommunikatsiyalaridan oqilona, ta'lim maqsadlarida foydalanishdir. Nafaqat, o'qituvchilarni balki, yosh ta'lim olib kelayotgan kadrlarni vaqtlarini tejash hamda, ishga, o'qishga bo'lgan layoqatlarini oshirishga yordam beradi. Qo'shimcha tarzda shuni ta'kidlash mumkinki, raqamlashtirish tizimi hozirda butun dunyo globallashtirishda keng joriy etilgan tashabbus sifatida qaralmoqda. Bunda- ta'lim oluvchi muntazam va mustaqil o'qib-o'rganishi, o'zi qaror qabul qilishi va tajribasi orqali fikr yuritishi so'ngra, ish ko'rmog'i lozim. Hozirgi kunda raqamli iqtisodiyotni rivojlanayotgani hechkimga sir emas. Masalan, biznes-tadbirkorlikni qo'llab-quvvatlashda raqamli iqtisodiyotni kirib kelishi, bu sohani rivojiga va xalqaro hamkorli va pul almashinuvi, oldi-sotdi yo'nalishlarini osonlashtirib, online tarzda yangi hamkorlik, fikr almashinuvi va tajribalarni bo'lishish hamda uni nazorat qilishni osonlashtiradi. Bank sohasi, ishlab chiqarish va kompyuter IT sohasi, bu esa yangi ish o'rinlarini va yangi dizayn sohaslarini rivojiga o'zi katta hissasini qo'yadi.

Mamlakatimizda mustaqillikdan so'ng xorijiy tillarni mukammal o'zlashtirish, o'qituvchilarning zamonaviy o'qitish tizimini rivojlantirish hamda ularga mos ravishda o'quv qo'llanma-darsliklarni yaratish va Jahon ta'lim tizimi me'zonlariga moslashtirish, amalda tatbiq etish konsepsiyalari bir nechta boshqichlarga bo'lingan.

O'zbekiston Respublikasi Prezidentining 2012-yil 10-dekabrda PQ-1875-sonli Qarorini qabul qilinishi bilan undagi qaror amalda o'z kuchini topganini ko'rishimiz mumkin, unda-“ chet tillarini o'rganish tizimini takomillashtirish, zamonaviy pedagogik va axborot-kommunikatsiya texnologiyalardan foydalanish”. Shuningdek, chet tillariga o'qitishning uzluksiz va murakkab tizimini ishlab chiqish, o'qimishli va salohiyatli, zamonaviy fikrlovchi, el-yurt ravnaqiga o'z hissasini qo'sha oladigan yosh kelajak vorislarini shakllantirish kabi, g'oyalari ilgari surilgan. 2021-yil 19-mayda “O'zbekiston Respublikasida xorijiy tillarni o'rganish va o'rgatishni ommalashtirish va ularni qo'llab-quvvatlash” haqida PQ 3775-sonli qarorda tasdiq etilgani ya'na, bir so'zimizni isbotidir. Raqamli iqtisodiyot bilan chet tillarini o'rganish va o'rgatish, paralel olib borish va uni ilm-fan rivojiga qo'ygan O'zbekistondagi OTMlarini o'qitishga qo'yilgan asosiy maqsadlardan biri- talaba yoshlarni teran fikrlovchi, zukko har tomonlama yetuklikka erishuvchi va jamiyatdagi xalqaro butun jahon yangiliklarni, axborot texnologiyalari orqali yangi dastur-metodik bilimlarini o'rganish, o'rgatish aholini solohiyatini oshirish butun, jahon bilan hamjihatlikda yasash g'oyalari ham ilgari surilgan.

Raqamli ta'lim muhitining ahamiyati-Raqamli platformalar orqali ta'lim jarayonini tashkil qilish va individual ta'lim yondashuv va axborot-texnologiyalari, interaktiv ta'lim resurslari yordamida bilimlarni chuqurroq o'zlashtirish. Asosiy g'oyalardan biri esa- kasbiy kompetensiyalarni rivojlantirish, teoretik bilimlarni, amaliy ko'nikmalarni va muammolarni hal qilish ko'nikmalarini rivojlantirishdir. Bunda asosan, to'rt metodikani takomillashtirish zarurati paydo bo'ladi. Bular, shaxsga yo'naltirilgan ta'lim, interaktiv ta'lim texnologiyasi, kooperativ o'qitish metodikasi va nazorat, baholash tizimlarini o'zi ishiga oladi. Bu raqamli texnologiyalar bilan real jarayonlar va eksperimentlarni kompyuter texnologiyalar bilan sistemlash, kodlash va raqamlashtirish bu insoniyat hayotidagi qiyinchiliklar osonlashtirishga olib keladi. Raqamli tizim ishchilarga foydali tarafi ularni qilishi kerak bo'lgan turli qo'lda bajarilishi uchun ko'p vaqt olinadigan qog'oz, hujjat ishlaridagi qo'shimcha besamar vaqt o'tkazish ishlarini oldini oladi. Bu kelajakda tez orada amalga oshirilishi mumkin bo'lgan istiqbolli, amalda o'sib borayotgan sohalardan biridir.

Xulosa

Bugungi tez sur'atlarda rivojlanib borayotgan texnologiyalar asrida, axborot-texnologiyalarida foydalanish bu-butun dunyo bilan o'zaro kooperativ hamkorlik qilish, talaba yoshlarni esa, kognitiv qobiliyat salohiyatini, ijodkorlik va maqsadga erishish yo'lida boshib o'tiladigan bir nechta amaliy-tajribiy ishlarni rivojlantirishga va ular ustida ishlash va mavjud muammolarni ko'ra olish va baholash, ularga to'g'ri qarorlar chiqarish va hal qilish qobiliyatini takomillashtiradi.

Foydalanilgan adabiyotlar:

1. Oliy ta'lim tizimini raqamli avlodga moslashtirish konsepsiyasi. Evropa ittifoqi Erasmus+dasturining kōmagida//O'zbekiston Respublikasi Oliy va o'rta maxsus ta'lim vazirligi.-2019.-B.13.
2. Wang Q., H.L. Woo H.L., Quek C.L., Yang Y. and Liu M. Using the Facebook group as a learning management system: An exploratory study. Br. J. Educ. Technol, 2012, 43(3): 428-438.
3. O'zbekiston Respublikasida xorijiy tillarni o'rganishni ommalashtirish faoliyatini sifat jihatidan yangi bosqichga olib chiqish chora-tadbirlari to'g'risidagi O'zbekiston Respublikasining PQ-5117-son qarori. 2021-yil 19-may.
4. "Ta'lim, fan va innovatsiya" ma'naviy-ma'rifiy, ilmiy-uslubiy jurnal 2023-yil 2-son
5. 2023-yil 25-oktabr kuni T.N.Qori Niyoziy nomidagi O'zbekiston pedagogika fanlari ilmiy tadqiqot institutida o'tkazilgan "Raqamli ta'limning zamonaviy tendentsiyalari va ularni ta'lim-tarbiya jarayoniga tatbiq qilish yo'llari" mavzusidagi respublika ilmiy-amaliy anjuman materiallari to'plami.

O‘ZBEK VA TATAR MILLIY KORPUSLARI TOKSONOMIYASI

Muxamadkulova Muxlisa Baxtiyor qizi

O‘zMU Jurnalistka va o‘zbek filologiyasi fakulteti
Filologiya va tillarni o‘qitish: o‘zbek tili yo‘nalishi talabasi
E-mail: muxlisa1525@gmail.com

Ilmiy rahbar: **A.R. Iskandarova**
f.f.n., dotsent

Annotatsiya: Ushbu maqolada o‘zbek va tatar milliy korpuslari toksonomiyasi haqida so‘z boradi. Ikki qardosh tilning korpuslaridagi o‘xshash va farqli jihatlari, bu borada olib borilayotgan tadqiqotlar haqida nazariy xulosalar chiqariladi.

Kalit so‘zlar: korpus, lingvistik lug‘atlar, web sahifalar, morfologik ma’lumotlar bazasi, multimedia, dialect, audio.

Аннотация: В данной статье говорится о систематике узбекско-татарского национального корпуса, сделаны теоретические выводы о сходстве и различиях корпусов двух родственных языков, а также проведены исследования в этом направлении.

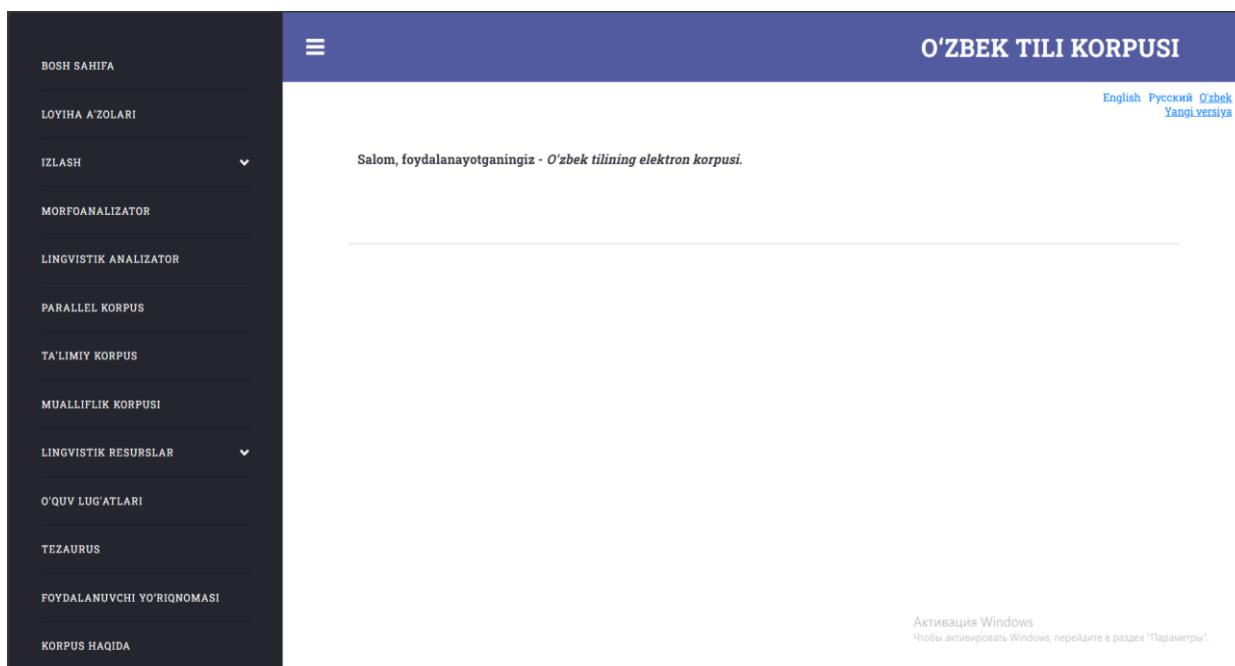
Ключевые слова: корпус, лингвистические словари, веб-страницы, морфологическая база данных, мультимедиа, диалект, аудио.

Abstract: This article discusses the taxonomy of the Uzbek-Tatar national corpus. Theoretical conclusions are drawn about the similarities and differences in the corpora of the two sister languages, as well as research conducted in this area.

Keywords: corpus, linguistic dictionaries, web pages, morphological database, multimedia, dialect, audio.

Korpus (Corpus, pl. corpora) – lotin tilidan olingan bo‘lib, elektron ma’lumotlar bazasiga yig‘ilgan matnlar to‘plamidir. Korpus uchun arxiv yoki boshqa katta hajmli matnlardan muayyan bir tilning farqlarini, undagi janrlarni hamda me‘yoriy qoidalarni aks ettirish uchun foydalaniladi. Korpuslar, odatda, so‘z yoki nutq bilan bog‘liq xususiyatlarni aniqlovchi qo‘shimcha meta-ma’lumotlarga ega bo‘ladi. Bu meta-ma’lumotlar matn janri, muallifi, nashr etilgan vaqti va joyi haqida ma’lumot beradi [Abduraxmonova N., 2024; 25]. Korpus lingvistikasi zamonaviy amaliy tilshunoslikning ustuvor yo‘nalishi hisoblanadi. Bugungi kunda zamonaviy kompyuter texnologiyalaridan foydalangan holda lingvistik ma’lumotlar korpusini yaratish mutaxassislar oldida turgan vazifalardan biridir. Ushbu maqolada o‘zbek va tatar tillari korpuslarining toksonimiyasi tahlil qilindi.

Jahon ilm-fanida dastlabki yozma korpuslar sifatida Brown (1964), LOB (1978), FROWN (1999), FLOB (1998), Kolhapur (1978), ACE (1986), og‘zaki korpuslar sifatida SEU, LLC, SEC (ingliz tilining og‘zaki matn korpusi) korpuslari e’tirof etiladi [Abduraxmonova N., 2024]. Ushbu korpuslarning hajmi 1 milliondan ortiq so‘zni o‘z ichiga oladi. Korpus lingvistikasi sohasida Respublikamizda bir qator oliy ta’lim muassasalari, shuningdek, ilmiy tadqiqot institutlarida ilmiy izlanishlar olib borilmoqda. O‘zbek korpus lingvistikasida B.Mengliyev, N.Abduraxmonova, Sh.Shahobiddinova, Z.Xolmanova, S.Karimov, L.Raupova, Sh.Hamroyeva, G.Toirova, J.Djumabayeva, G.Ergasheva, A.Eshmo‘minov, Sh.Gulyamovalarning tadqiqotlarini qayd etish o‘rinlidir [Abduraxmonova N., 2024]. Ayni paytda o‘zbek tili korpusini yaratishga doir tadqiqotlardan biri professor Nilufar Abduraxmonova tomonidan olib borilmoqda. Tadqiqotning amaliy natijasi bilan <https://uzbekcorpus.uz/> platformasida tanishish mumkin.



Ushbu korpusda o'zbek tilida yaratilgan lingvistik lug'atlar, web sahifalar, o'zbek tilining morfologik ma'lumotlar bazasi, o'quv adabiyotlar hamda turli janrdagi ilmiy, rasmiy va badiiy uslubdagi matnlar majmuasi jamlangan. Ushbu korpusning tasnifi quyidagicha:

- *maqsadiga ko'ra: umumiy korpus;*
- *mavjud subkorpuslar: ko'p tilli parallel korpus, mualliflik korpusi, ta'limiy korpus, gazeta korpusi, bolalar nutqi korpusi, audio korpusi, multimedia korpusi, dialekt korpusi;*
- *mavjud lingvistik resurslar: o'zbek tili tezavrusi, ingizcha-o'zbekcha fe'lli frazemalar lug'ati.*

O'zbek tili korpusi <https://uzbekcorpus.uz/> bu o'zida 10 million so'zni o'zida jamlagan. O'zbek korpusida quyidagi lug'atlar ham o'rin olgan:

Lingvistik analizator (lingvistik tahlil o'tkazish instrumenti);

Tezavrus;

O'quv lug'atlari;

Lingvistik resurslar (sinonimlar lug'ati, terminologik ma'lumotlar bazasi).

O'zbek tili korpusining asosiy manbasi o'zbek, rus va ingliz tillari berilgan. Parallel korpus orqali boshqa tillardagi ma'lumotlar ham mavjud, xorijiy tillar kesimidagi parallel korpus kontentlari joylanmoqda.

Tatar tili korpusini yaratish ishlari 2010-yildan boshlangan bo'lib, tadqiqot natijalari bilan <https://www.corpus.tatar> platformasida tanishish mumkin. Korpus 500 milliondan ortiq so'zdan (>620 million token) iborat. Tatar tilining yozma korpusini yaratish bo'yicha izlanishlar 2010-yildan boshlangan.

Yozma korpusga kiritilgan matnlar asosan uchta uslubga tegishli: jurnalistika ($\approx 60\%$), badiiy adabiyot ($\approx 35\%$) va gumanitar fanlar bo'yicha ilmiy adabiyotlar ($\approx 5\%$). Hozirgi vaqtda tatar tili korpusiga kiritish uchun material to'plash va uni qayta ishlash davom etmoqda. Tatar tili elektron korpusining asosiy maqsadi tatar tilining lug'at boyligi bo'yicha ilmiy tadqiqotlarni osonlashtirishdir. Bundan tashqari, yozma korpus til o'qitishda, turli uslubdagi hujjatlarni tuzishda ma'lumotnoma sifatida ahamiyatga ega.

Tatar tili korpusining asosiy menyusi quyidagicha:

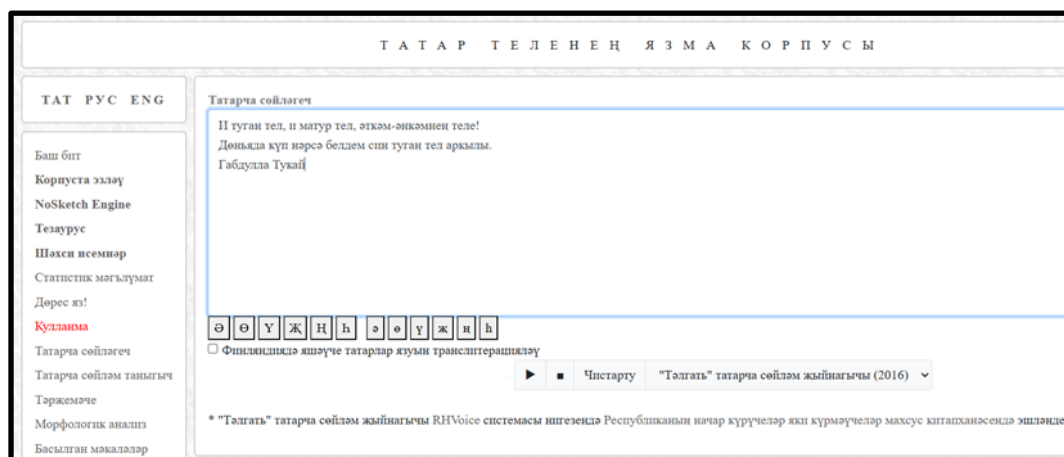


Tatar tili korpusini yaratishda Mansur Sayxunov, Rustem Xasenov va boshqa olimlar ishtirok etgan. Tatar tilining yozma korpusi bo'yicha barcha ishlar faqat loyiha ishtirokchilarining bo'sh vaqtida amalga oshiriladi. Shunday bo'lsa ham tatar korpusi o'z ichiga 500 mln dan ortiq so'zni qamrab olgan. Tatar tili mutaxassislari tomonidan tatarcha yozma matnlarni korpus ostiga jamlash [<https://www.corpus.tatar>] orqali ona tilining rivojlanishiga, yangi texnologiyalar asosida taraqqiy etishiga hissa qo'shmoqda.

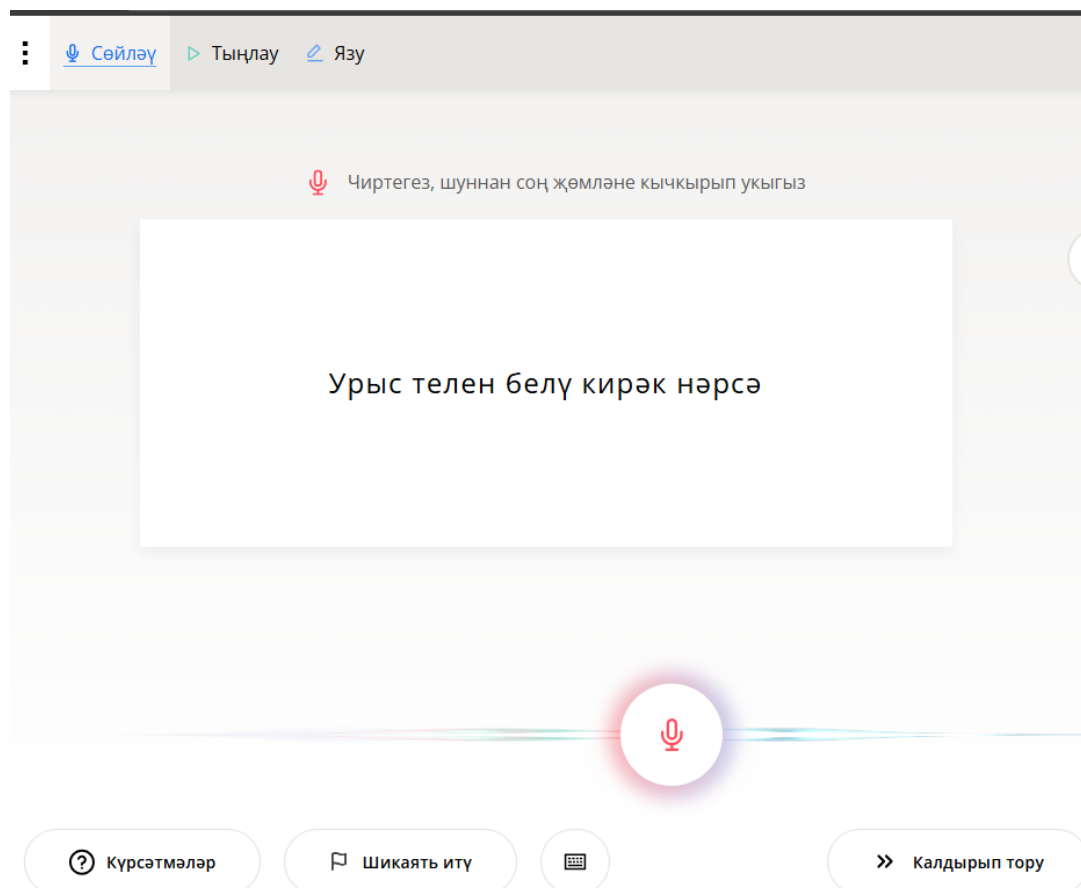
Tatar tili korpusida quyidagi bo'limlar mavjud:

- *tezaurus*;
- *shaxsiy ismlar*;
- *statistik ma'lumotlar*;
- *imlo tekshiruv*;
- *tatarcha matndan nutqqa*;
- *tatarcha nutqdan matnga*;
- *nashr etilgan maqolalar*;
- *morfologik analiz*;
- *tarjima*.

“**Tatarcha matndan nutqqa**” bo'limiga kirilganda, istalgan matnning tatarcha audio variantini eshitish mumkin. Masalan:

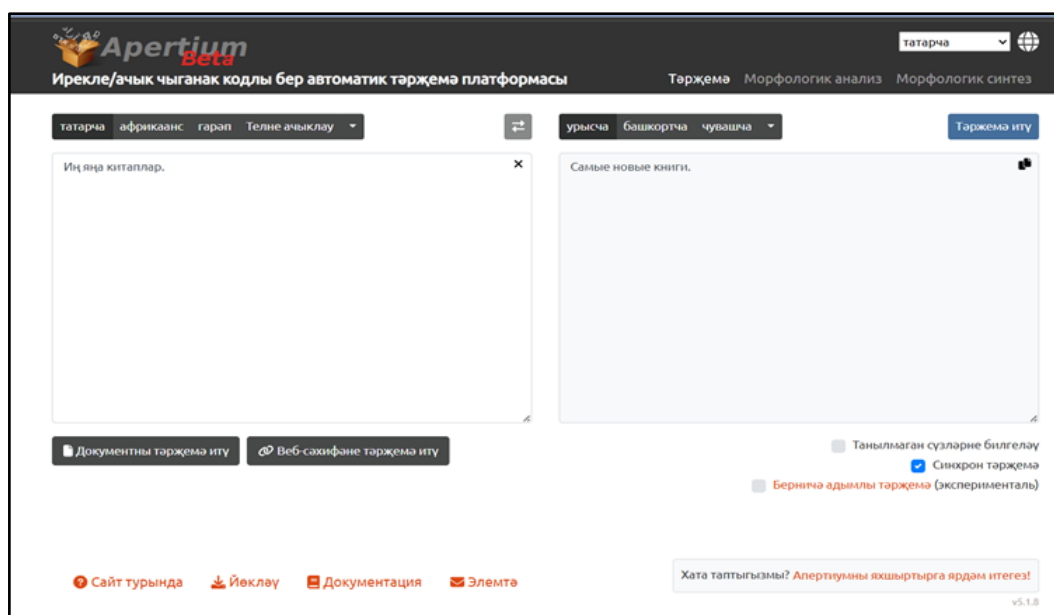


“Tatarcha nutqdan matnga” bo‘limi to‘g‘ridan-to‘g‘ri <https://commonvoice.mozilla.org/tt> saytiga ulangan bo‘lib, ko‘ngillilar tomonidan to‘plangan ma’lumotlar bazasi orqali ishlaydi. Buni quyidagi misolda ko‘rish mumkin:



“Tarjima” bo‘limi so‘zlarni bevosita tarjima qilish imkonini beradi. Bu bo‘limdan 170 dan ortiq tildagi matnlarni tarjima qilish mumkin. Bu tillar ichida o‘zbek tili ham bor.

“Morfologik analiz” va “Morfologik sintez” bo‘limlarida, bevosita, so‘zlarni morfologik tahlil qilish mumkin:



Yuqoridagi “Tarjima”, “Morfologik analiz” va “Morfologik sintez” bo‘limlari <https://beta.apertium.org/> saytiga ulangan. Tatar korpusi ko‘plab saytlar bilan hamkorlik qiladi va korpusning “Foydali ma’lumotlar” bo‘limida qaysi saytlardan foydalanilgani haqida ma’lumot berib o‘tadi.

Korpus lingvistika yildan-yilga rivojlanmoqda. Xususan, o‘zbek va tatar korpuslari ham yangi manbalar bilan boyitilmoqda. Yuqorida tahlil qilingan har ikkala korpusdan lingvistik tahlil qilish, avtomatlashtirilgan tarjima tizimlarini rivojlantirish va sun’iy intellektga asoslangan til modellarini yaratishda keng qo‘llaniladi. Bu korpuslar tilshunoslarga keng hajmli matnlar asosida tadqiqot olib borish imkonini beradi. Qolaversa, fonetika, morfologiya va sintaksis bo‘yicha aniq natijalarga erishishga yordam beradi.

Adabiyotlar ro‘yxati:

1. Abdurahmonova N. Korpus lingvistikasi. T.: GlobeEdit, 2024. –B. 450.
2. <https://uzbekcorpus.uz/>
3. <https://www.corpus.tatar>
4. Martin W. Practical Corpus Linguistics An Introduction to Corpus-Based Language Analysis. Oxford, 2016. – P.16.
5. Sandra K., Heike Z. Corpus linguistics and linguistically annotated corpora. – New York: Bloomsbury Academic, 2015. –P. 4.
6. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: [uzbekcorpus. uz](https://uzbekcorpus.uz). In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.

NUTQ SINTEZATORI UCHUN MO'LJALLANGAN NLP PAKETLARI

O'rinboyeva Ro'za Toliboyevna

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti magistranti

E-mail: orinboyevaroza@gmail.com

To'rayeva Umida Toxirjon qizi

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti magistranti

E-mail: umidasng@gmail.com

Annotatsiya. Ushbu maqola nutq sintezatorlari uchun mo'ljallangan tabiiy tilni qayta ishlash (NLP) paketlarini ko'rib chiqadi. Nutq sintezatorlari matnni nutqqa aylantirish uchun ishlatiladi va NLP paketlari bu jarayonni yaxshilashda muhim rol o'ynaydi. Maqolada NLTK, spaCy, Transformers, Stanford CoreNLP, Gensim kabi NLP paketlarining xususiyatlari, afzalliklari va kamchiliklari muhokama qilinadi. Shuningdek, nutq sintezatorlari uchun NLP paketlarini tanlashda e'tiborga olish kerak bo'lgan omillar ham ko'rib chiqiladi.

Kalit so'zlar. Nutq, model, sintezator, qayta ishlash.

Abstract. This article reviews natural language processing (NLP) packages for speech synthesizers. Speech synthesizers are used to convert text into speech, and NLP packages play an important role in improving this process. The article discusses the features, advantages, and disadvantages of NLP packages such as NLTK, spaCy, Transformers, Stanford CoreNLP, Gensim. It also discusses factors to consider when choosing NLP packages for speech synthesizers.

Keywords. Speech, model, synthesizer, processing.

Аннотация. В данной статье рассматриваются пакеты обработки естественного языка (NLP), разработанные для синтезаторов речи. Синтезаторы речи используются для преобразования текста в речь, и пакеты обработки естественного языка играют важную роль в совершенствовании этого процесса. В статье рассматриваются особенности, преимущества и недостатки таких пакетов обработки естественного языка, как NLTK, spaCy, Transformers, Stanford CoreNLP, Gensim. В нем также обсуждаются факторы, которые следует учитывать при выборе пакетов обработки естественного языка для синтезаторов речи.

Ключевые слова. Речь, модель, синтезатор, обработка.

Nutq sintezatorlari turli xil ilovalarda, jumladan, virtual yordamchilar, navigatsiya tizimlari va ta'lim dasturlarida keng qo'llaniladi. Ushbu ilovalarda nutq sintezatorlari tabiiy va tushunarli nutqni yaratishi kerak. Tabiiy tilni qayta ishlash (NLP) paketlari nutq sintezatorlariga matnni tahlil qilish va tabiiy nutqni yaratishda yordam beradi. Tabiiy tilni qayta ishlash jarayonida turli vazifalarni bajarishga mo'ljallangan dasturiy ta'minotlardan foydalaniladi. Ular NLP paketlari deb nomlanadi.

Bugungi kunda jahon olimlari tomonidan tabiiy tilni qayta ishlash jarayoni uchun **NLTK (Natural Language Toolkit)**, **spaCy**, **Transformers (Hugging Face)**, **Stanford CoreNLP**, **Gensim** kabi paketlardan foydalanilmoqda.

1. NLTK (Natural Language Toolkit): NLTK – bu inson tilidagi ma'lumotlar bilan ishlashga mo'ljallangan Python dasturlash tilida yaratilgan platforma. Uning tarkibida korpus va leksik resurslar uchun mo'ljallangan interfeyslar mavjud bo'lib, ular matnlarni tasniflash, tokenlash, stemming, teglash hamda tahlil qilish kabi vazifalarni bajaradi. Matnni semantik qayta ishlash uchun kutubxonalar to'plamini taqdim etadi. Bu ochiq manbali, bepul platforma hisoblanib, talabalar, o'qituvchilar, tilshunoslar, turli soha tadqiqotchilari uchun ham mos keladi [NLTK :: Natural Language Toolkit]. **Steven Bird, Ewan Klein, and Edward Loper**lar

tomonidan maxsus elektron kitob yaratilgan. Unda Python dasturlash tili orqali matnni qayta ishlash haqida batafsil yoritilgan [NLTK Book]:

Natural Language Processing with Python

– Analyzing Text with the Natural Language Toolkit

Steven Bird, Ewan Klein, and Edward Loper

This version of the NLTK book is updated for Python 3 and NLTK 3. The first edition of the book, published by O'Reilly, is available at http://nltk.org/book_1ed/. (There are currently no plans for a second edition of the book.)

0. [Preface](#)
1. [Language Processing and Python](#)
2. [Accessing Text Corpora and Lexical Resources](#)
3. [Processing Raw Text](#)
4. [Writing Structured Programs](#)
5. [Categorizing and Tagging Words](#) (minor fixes still required)
6. [Learning to Classify Text](#)
7. [Extracting Information from Text](#)
8. [Analyzing Sentence Structure](#)
9. [Building Feature Based Grammars](#)
10. [Analyzing the Meaning of Sentences](#) (minor fixes still required)
11. [Managing Linguistic Data](#) (minor fixes still required)
12. [Afterword: Facing the Language Challenge](#)

[Bibliography](#)

[Term Index](#)

2. **spaCy** – katta hajmdagi matnlar bilan ishlashga mo'ljallangan Python kutubxonasi. Dasturlash tillari bo'yicha mutaxassis Yuli Vasiliyevning “*Natural Language Processing with Python and spaCy*” nomli kitobida spaCy kutubxonasi va uning matnni qayta ishlashdagi mavjud imkoniyatlari batafsil ochib bergan [Vasiliyev Y., 2020].

3. **Transformers (Hugging Face)**. Ushbu ochiq kodli kutubxona aynan katta tilli modellarni (LLM) ishlatish uchun mo'ljallangan. U nutq sintezida tabiiy nutqni yaratish uchun ishlatiladigan oldindan tayyorlangan modellarning keng doirasini taklif etadi. Ushbu kutubxona yordamida matnni turli toifalarga ajratish, matndan yangi matn yaratish, matn asosida berilgan savollarga javob berish, bir tildan boshqa tilga tarjima qilish hamda uzun matndan qisqa xulosa olish mumkin. Transformers kutubxonasi BERT, GPT, T5 eng mukammal o'qitilgan modellarni o'z ichiga oladi. Lyuis Tunstall, Leandro fon Verra va Tomas Volf tomonidan yozilgan “*Natural Language Processing with Transformers*” nomli kitobda ushbu kutubxona imkoniyatlari, undan foydalanish usullari haqida keltirilgan [Tunstall L., Werra L., Wolf T., 2022].

Ushbu paketlar kompyuterga inson tilini tushunish, uni tahlil qilish va generatsiya qilish imkoniyatini beruvchi yordamchi vosita hisoblanadi. Manbalarda nima maqsadda ishlatilishiga ko'ra quyidagicha farqlanadi:

- kiritilayotgan matnga oldindan ishlov berish;
- ortiqcha belgilarni olib tashlash;
- so'zlarni ajratib berish;
- kiritilayotgan matnning qaysi tilda ekanligini aniqlash;
- berilgan matndagi so'zlarni turkumga ajratish;
- matndagi nomlarni aniqlash (masalan, shaxs nomi, joy nomi, tashkilot nomi kabi);
- matnni sentiment tahlil qilish (ijobiy yoki salbiy);
- xulosalash;
- matn asosida savollarga javob berish;
- inson tili kabi tabiiy matn yaratish.

Har bir vazifa uchun mo'ljallangan alohida paketlar – dasturiy ta'minotlar mavjud bo'lib, ushbu faslda nutq sintezatori uchun yaratilgan va bugungi kunda eng ko'p foydalanilayotgan NLP paketlari tahlil qilinadi.

Nutq sintezatori yaratish uchun nafaqat NLP paketlari, balki til modellari ham muhim ahamiyatga ega. Ular matnni tabiiy va tushunarli nutqqa aylantirishga mo'ljallangan bo'lib, nutqqa xos sifatlarini o'zida jamlaydi va quyidagi vazifalarni bajaradi:

Matnni tahlil qilishda:

- Matnni fonetik birliklarga (fonemalar) ajratadi;
- So'zlarning talaffuzini aniqlaydi;
- Intonatsiya va urg'uni aniqlaydi;
- Gapning sintaktik tuzilishini tahlil qiladi.

Nutqni generatsiya qilishda:

- Fonemalarni nutq signallariga aylantiradi;
- Tabiiy va tushunarli nutqni yaratadi;
- Turli xil ovozlari va nutq uslublarini yaratadi.

Nutqqa ishlov berish jarayonida til modellari talaffuzning to'g'riligi hamda aniqligini oshirib beradi. Nutqning tabiiy va tushunarli bo'lishi ta'minlaydi. Til modellarining qamrovi qanchalik keng bo'lsa, turli xil ovoz hamda uslubda berilgan nutqni tahlil qila oladi. Barcha tildagi nutqni qabul qilish imkoniyatiga ega bo'ladi. Til modellarining sifati nutqqa qanday tezlikda ishlov berishiga ko'ra belgilanadi. Nutq sintezatori uchun qo'llaniladigan til modellarini quyidagicha tasniflash mumkin:

- *Fonetik modellar:* Matnni fonemalarga aylantirish uchun ishlatiladi;
- *Akustik modellar:* Fonemalarni nutq signallariga aylantirish uchun ishlatiladi;
- *Duratsion modellar:* Nutqning davomiyligini aniqlash uchun ishlatiladi;
- *Intonatsion modellar:* Nutqning ohangini aniqlash uchun ishlatiladi;
- *Neyron tarmoqlar:* Yuqori sifatli va tabiiy nutqni yaratish uchun ishlatiladi. Zamonaviy nutq sintezatorlari uchun, ko'p hollarda, WaveNet, Tacotron, GPT, BERT kabi modellar foydalaniladi.

O'zbek nutq sintezatori yaratish o'zbek tilining o'ziga xos xususiyatlari tufayli murakkab vazifa hisoblanadi. o'zbek tilining aglyutinativligi, singarmonizmning saqlanganligi hamda murakkab morfologik tuzilishi nutq sintezatori uchun maxsus NLP paketlarini talab qiladi. Yuqorida sanab o'tilgan NLP paketlari o'zbek nutq sintezatori yaratishda yordam berishi mumkin bo'lsa-da, ularning aksariyati o'zbek tili uchun optimallashtirilmagan. Ayni paytda kompyuter lingvistlari uchun quyidagi yo'nalishlarda tadqiqot olib borish vazifasi yuklanadi:

- o'zbek tili uchun maxsus NLP paketlarini ishlab chiqish;
- o'zbek tilining nutq korpusini yaratish;
- o'zbek tilining talaffuz lug'atini yaratish;
- o'zbek tilining intonatsiya modelini yaratish.

Ushbu tadqiqotlar natijasida o'zbek tiliga moslashgan, yuqori sifatli nutq sintezatori yaratish mumkin. Bu o'zbek tilida so'zlashuvchi aholi uchun turli xil ilovalar, jumladan, virtual yordamchilar, navigatsiya tizimlari va ta'lim dasturlarida foydalanish imkoniyatini yaratadi.

Foydalanilgan adabiyotlar ro'yxati:

1. Vasiliev Y. Natural Language Processing with Python and spaCy. San Francisco, 2020. P. 196. Natural Language Processing with Python and spaCy: A Practical Introduction - Yuli Vasiliev - Google Books
2. Tunstall L., Werra L., Wolf T. Natural Language Processing with Transformers. 2022. Natural Language Processing with Transformers - Lewis Tunstall, Leandro von Werra, Thomas Wolf - Google Books
3. NLTK :: Natural Language Toolkit
4. NLTK Book

5. Indig B., Simonyi A., Ligeti-Nagy N. What's wrong, python? – A visual differ and graph library for NLP in python // 11th International Conference on Language Resources and Evaluation. LREC – 2018.
6. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
7. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
8. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzworonet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
9. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus. uz. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.

KORPUS MATNLARINI TEMATIK MODELLASHTIRISHDAGI ZAMONAVIY YONDASHUVLAR

Aloyev Narzillo Raxmatilloevich

Alisher Navoiy nomidagi

Toshkent davlat o'zbek tili va adabiyoti universiteti

tayanch doktorant

E-mail: vip.alayev@gmail.com

Annotatsiya: Tabiiy tilni qayta ishlash (Natural language processing, NLP), asosan, matnni qayta ishlashda qo'llaniladi va NLP yordamida chatbotlar, avtomatik tahrir va tahlil, nutqni tanib olish, avtomatik tarjima, ijtimoiy tarmoq monitoringi va elektron pochta filtrlash kabi ilovalar ishlab chiqiladi. NLPda tematik modellashtirish – bu katta hajmdagi matnlarni avtomatik ravishda umumlashtirish uchun ishlatilishi mumkin bo'lgan algoritmlar to'plami. Matn xususiyatlari soni juda ko'p bo'lgani sababli modellarni o'qitish qiyinlashadi va modellarning samaradorligini pasaytiradi. Bugungi kunda til korpusi matnlarida tematik modellashtirish va o'lchovni qisqartirishni turli xil algoritmlar bilan amalga oshirish mumkin: LDA, Non-negative Matrix Factorization, LSA, PLSA, Lda2Vec, Bert Topic-BERT. Ushbu maqolada tematik modellashtirishning tavsifi, qo'llanish sohalari, usullari, yondashuvlari haqida ma'lumot keltiriladi va LDA usuli orqali til korpusi matnlarini intellektual qayta ishlash orqali mavzular aniqlanadi va t-SNE metodi vositasida vizualizatsiya qilinadi.

Kalit so'zlar: Tematik modellashtirish, mashinali o'qitish, Yashirin Dirixle tahlili (LDA), til korpusi, mavzular tasnifi.

Аннотация: Обработка естественного языка (NLP) в основном используется при обработке текста, следовательно используется для разработки таких приложений, как чат-боты, автоматическое редактирование и анализ, распознавание речи, автоматический перевод, мониторинг социальных сетей и фильтрация электронной почты. В NLP тематическое моделирование — это набор алгоритмов, которые можно использовать для автоматического обобщения больших объемов текста. Из-за большого количества текстовых функций усложняется обучение моделей и снижается производительность моделей. Сегодня тематическое моделирование и уменьшение размерности текстов языкового корпуса можно выполнять с помощью различных алгоритмов таких как: LDA, Non-negative Matrix Factorization, LSA, PLSA, Lda2Vec, Bert Topic-BERT. В данной статье представлена информация по описанию тематического моделирования, областей применения, методов и подходов, а темы идентифицируются путем интеллектуальной обработки текстов языкового корпуса методом LDA и визуализируются с помощью метода t-SNE.

Ключевые слова: Тематическое моделирование, машинное обучение, латентный анализ Дирихле (LDA), лингвистический корпус, классификация тем.

Annotation: Natural language processing (NLP) is mainly used in text processing and used to develop applications such as chatbots, automatic editing and analysis, speech recognition, automatic translation, social network monitoring and email filtering. In NLP, topic modeling is a set of algorithms that can be used to automatically summarize large amounts of text. Due to the large number of text features, it becomes difficult to train the models and reduces the performance of the models. Today, topic modeling and dimension reduction in language corpus texts can be performed with various algorithms such as: LDA, Non-negative Matrix Factorization, LSA, PLSA, Lda2Vec, Bert Topic-BERT. This article provides information on the description of topic modeling, areas of application, methods, and approaches, and topics are identified by intellectual

processing of language corpus texts through the LDA method and visualized by means of the t-SNE method.

Keywords: Topic Modeling, Machine Learning, Latent Dirichlet Analysis (LDA), Linguistic Corpus, Topic Classification.

Kirish

Tematik modellashtirish – til korpusi matn (hujjat)laridagi mavzulardan soʻzlarni ajratib olishga qaratilgan NLP usuli. Matndan soʻzlarni ajratib olish koʻproq vaqt talab etadi va bu jarayon matndagi berib oʻtilgan mavzulardan ajratib olishdan koʻra ancha murakkabroqdir. Masalan, har bir 1000 ta matndan va har bir matnda oʻrtacha 500 tadan soʻz mavjud til korpusini tahlil qilamiz. Demak, ushbu korpusni qayta ishlash uchun $500 \times 1000 = 500000$ ta amalni bajarish kerak boʻladi. Shunday qilib, maʼlum mavzularni oʻz ichiga olgan matnni tahlil qilishda, agar oʻrtacha 5 ta mavzudan iborat boʻlsa, ishlov berish 5×500 soʻz = 2500 ta *amal (thread)*ni tashkil qiladi. Bu butun bir hujjatni qayta ishlashdan koʻra soddaroq koʻrinadi, shuning uchun tematik modellashtirish shu kabi muammolarni hal qilishda qoʻl keladi va jarayonni vizualizatsiya qilishni osonlashtiradi.

NLPda matnni qayta ishlashni osonlashtiradigan matnga boshlangʻich ishlov berish bosqichlarini amalga oshirish lozim:

- *Nomuhim soʻzlar va tinish belgilarini olib tashlash;*
- *Stemming;*
- *Lemmatizatsiya;*
- *Countvectorizer yoki Tf-Idf usuli yordamida statistik qiymatlarni shakllantirish.*
- *Stemming va Lemmatizatsiya*

Oʻzbek tili korpusi matnlarini stemlash va lemmalash boʻyicha B.Elov va Z.Xusainovalar bir qator ilmiy tadqiqotlarni olib borganlar. Oʻzbek tilidagi soʻzlarning stemini aniqlashda oʻzak va qoʻshimchani bitta oʻzak bilan omonim boʻlishi, soʻzlarga qoʻshimcha qoʻshilganda tovush oʻzgarishiga uchrashi va neologizm hamda NERlarni stemmlash kabi muammolar yuzaga kelishi mumkinligini qayd etishgan va bu muammoni hal etish modellarini keltirishgan. Stemni aniqlashdagi tovush oʻzgarishiga uchrashi muammosini hal qilish uchun birinchi bosqichda oʻzak va qoʻshimchalarning chegaralari aniqlanib, ikkinchi bosqichda esa lemmatizatsiya amalga oshirilgan. Lemmatizatsiya natijasida xato hosil qilingan stemlar lugʻatda mavjud oʻzak (root)ga oʻzgartiriladi. Maqolada keltirilgan usullar B.Elov, R.Alayev, Sh.Hamroyeva va Z.Xusainova tomonidan ishlab chiqilgan oʻzbek tili morfologik analizatoriga tadbiiq etilgan.

Agglyutinatativ tillarda oʻzak va qoʻshimchalarni birlashtirib soʻz hosil qilinadi. Oʻzakka qoʻshimchalar qoʻshilganda fonetik uygʻunlik va disgarmoniya yuzaga kelgani uchun ham fonetik, ham morfologik oʻzgarishlarni tahlil qilish zarur. Koʻpgina NLP vazifalarni hal qilishda soʻz shakllarini ularning oʻzakkacha qisqartirish (stemlash)ga toʻgʻri keladi [5]. Soʻzdan barcha flektiv affikslarni olib tashlash va soʻzning qolgan qismini lemmatizatsiya qilish tabiiy tilni qayta ishlash (NLP)ning muhim vazifalaridan biri hisoblanib, ushbu jarayon **stemming** deb yuritiladi. Stemming jarayoni axborot qidirish (IR, Information Retrieval) tizimlarida muhim ahamiyat kasb etadi. Agglyutinatativ tillar uchun stemmingni amalga oshirish uchun baʼzi terminlar izohini keltiramiz:

Lemma (leksema) – faqat oʻzakdan yoki oʻzak+soʻz yasovchi qoʻshimcha shaklidan iborat boʻladi. **Leksema** (yun. lexis – soʻz, ifoda) – til qurilishining leksik maʼno anglatuvchi lugʻaviy birligi. Leksema bildiradigan maʼno soʻzning material qismi: maʼlum tovush kompleksini maʼlum obyektiv voqelikka bogʻlash bilan kishi ongida yuzaga keladigan mazmun-mundarija.

Stem – soʻzshaklning qoʻshimchalarini kesib tashlashdan hosil boʻluvchi qism boʻlib, baʼzi hollarda maʼno anglatmasligi mumkin. Shuningdek, stem soʻzning morfologik oʻzagi bilan aynan mos boʻlmashligi yoki mos tushishi mumkin.

Turk va uygʻur tillarida stemming jarayoniga quyidagicha taʼrif berilgan:

Stemming – bu soʻzga qoʻshilgan *derivatsion* va *flektiv* qoʻshimchalarni olib tashlash orqali uning oʻzagigacha qisqartirish vazifasidir.

Lemmatizatsiya – lemma(leksema)ning maʼnoli shaklini aniqlash va uni flektiv/hosila shakllari oʻrnida ishlatishni oʻz ichiga olgan tabiiy tilni qayta ishlash usuli.

Lemmatizatsiya va stemming – tabiiy tilni qayta ishlash (NLP)da tahlil murakkabligini kamaytirish maqsadida ishlatiladigan ikkita usul. Bu ikkala usul ham soʻzlarni asosiy shakllariga qisqartirishdan iborat, ammo ular bu amalni bajarish maqsadi va unga erishish usullari bilan farqlanadi.

Lemmatizatsiya – soʻzning asosiy shakli boʻlgan **lemmasi(leksema)ga** qisqarish jarayoni. Bu asosiy shakl koʻpincha soʻzning **lugʻatdagi shakli** deb ham ataladi. Masalan, “*bajargan*” soʻzining lemmasi “*bajarmoq*”, “*sening*” soʻzining lemmasi esa “*sen*” hisoblanadi. Lemmatizatsiya soʻzning *kontekstini*, jumladan, *POS tegini*, *zamon* va *turkumini* hisobga oladi.

Stemming – kontekstni hisobga olmagan holda soʻzning asosiy shakli boʻlgan oʻzagiga qisqartirish jarayoni. Stemlar koʻpincha soʻzning oʻzak shakli deb ataladi. Masalan, “*yugurish*” soʻzining oʻzagi “*yugur*”, “*sening*” soʻzining oʻzagi esa “*sen*”, *sening* soʻzining stemi esa “*se*”dir. *Stemming* – bu NLP masalalarini yechishdagi murakkabliklarni kamaytirishning asosiy omili boʻlib, soʻzning kontekstini hisobga olmaydi. Bu esa soʻz maʼnosini ifodalashda koʻplab xatolarga olib kelishi mumkin.

Lemmatizatsiya – qidiruv tizimi samaradorligini oshirish uchun soʻzlarni lemma (asosiy) shakliga qisqartirish jarayoni. Ushbu jarayon, ayniqsa, qidiruv natijalarining aniqligini va korpus matni kontentini tushunish uchun foydali boʻlishi mumkin. Lemmatizatsiyaning asosiy afzalliklaridan biri shundaki, u *qidiruv tizimlariga matn konteksti* va “*maʼnosi*”ni yaxshiroq tushunishga yordam beradi.

Oʻzbek tili leksik birliklarini stemlash va lemmalash boʻyicha fikr-mulohazalar B.Elov, Sh.Hamroyeva, R.Alayev va Z.Xusainova tomonidan olib borilgan ilmiy tadqiqotlarda keltirilgan.

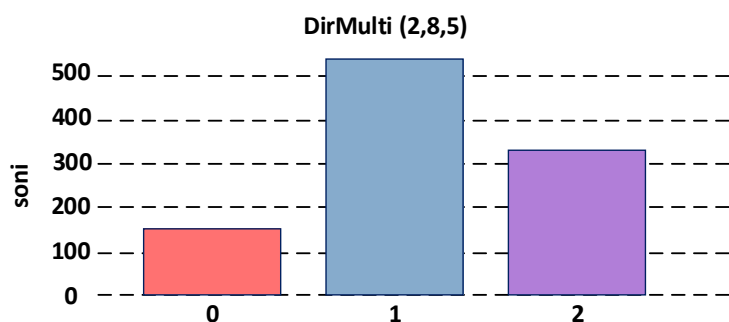
Bugungi kunda til korpusi matnlarini tematik modellashtirish asosan LDA (Latent Dirichlet Allocation) yordamida amalga oshirilmoqda. Tematik modellashtirish hujjatlar toʻplamini eng yaxshi tavsiflovchi mavzularni aniqlash vazifasini bajaradi. Bu mavzular faqat tematik modellashtirish jarayonida paydo boʻladi (shuning uchun ular **latent/yashirin** deb ataladi). Tematik modellashtirishning mashhur usullaridan biri Latent Dirichlet Taqsimoti (LDA) hisoblanadi.

Tematik modellashtirish – bu maʼlumotlarni turli qismlarga ajratuvchi klaster algoritmlari kabi shablonlarni topish orqali mavzularni aniqlash yoki ajratishga moʻljallangan nazoratsiz mashinali oʻqitish yondashuvidir. Tematik modellashtirishda soʻz klasterlaridan shablonlarni ajratish va hujjatdan soʻzlarning uchrash chastotasini aniqlab olish orqali amalga oshiriladi. Shunga asoslangan holda tematik modellashtirish hujjatni turli mavzularga ajratadi. Hujjatlar soni koʻp boʻlganda va ularda qanday turdagi maʼlumotlar mavjudligini aniqlash lozim boʻlganida, bunday modellashtirish juda foydali hisoblanadi.

Latent Dirichlet Allocation (LDA) usuli

LDA usulidagi **latent** – matnli maʼlumotlarda mavjud boʻlgan yashirin mavzularni anglatadi, **dirixle** – taqsimot shakli boʻlib hisoblanadi. Dirichlet taqsimoti odatdagi taqsimotdan farq qiladi. Matnli maʼlumotlarni qayta ishlashda ML algoritmlarini qoʻllash kerak boʻlganda, maʼlumotlar *normal taqsimlanishi* yoki *Gauss taqsimotiga* amal qilishi kerak boʻladi. Normal taqsimot maʼlumotlarni haqiqiy son formatida ifodalaydi. Dirichlet taqsimotida esa oʻz navbatida maʼlumotlar taqsimoti yigʻindisi 1 ga teng boʻladi.

Misol uchun,



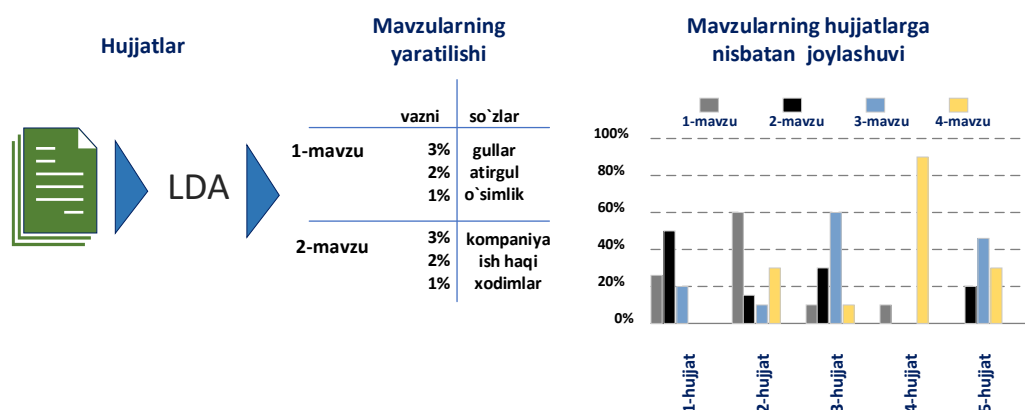
1-rasm. Dirixle taqsimotiga namuna

Normal taqsimot bizga ma'lumotlarning ma'nosidan kelib chiqqan holda og'ishini ko'rsatadi va berilgan ma'lumotlardagi tafovutga qarab o'zgaradi. Tafovut yuqori bo'lsa, ma'lumotlardagi qiymatlar ma'nosiga qaraganda kichik va kattaroq bo'ladi hamda assimetrik taqsimotni hosil qilishi mumkin. Agar tafovut past ko'rsatkichga ega bo'lsa, natijalar o'z navbatida ma'nosiga yaqin bo'ladi va agar tafovut nolga teng bo'lsa, ma'nosiga teng bo'ladi.

LDA usulidan foydalangan holda tematik modellashni amalga oshirish

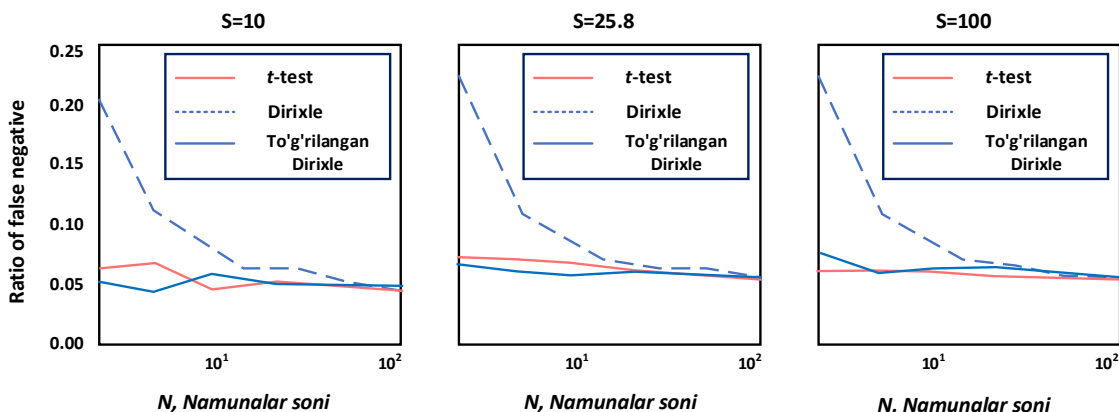
```
# Grid Search yordamida parametrlarni sozlash
from sklearn.model_selection import GridSearchCV
from sklearn.decomposition import LatentDirichletAllocation
from sklearn.manifold import TSNE
grid_params = {'n_components': list(range(5,10))}
# LDA modeli
lda = LatentDirichletAllocation()
lda_model = GridSearchCV(lda,param_grid=grid_params)
lda_model.fit(document_term_matrix)
# LDA modeli uchun hisoblagichlar
lda_model1 = lda_model.best_estimator_
print("LDA modelining eng yaxshi parametrlari", lda_model.best_params_)
print("LDA modeli uchun eng yaxshi ehtimol ballari ",lda_model.best_score_)
print("LDA modeli o'quv ma'lumotlaridagi noaniqlik",
lda_model1.perplexity(document_term_matrix))
```

LDA uchta muhim giperparametrga ega. Birinchisi **“alpha”** – hujjat-mavzu zichlik nisbatini ifodalaydi, ikkinchisi **“beta”** – mavzudagi so'zlarning zichligini bildiradi, uchinchi **“k”** – hujjat klasterlanishi yoki bo'linishi kerak bo'lgan mavzular sonini ifodalovchi komponentlar soni.



2-rasm. Korpus matn(hujjat)larini tematik modellash

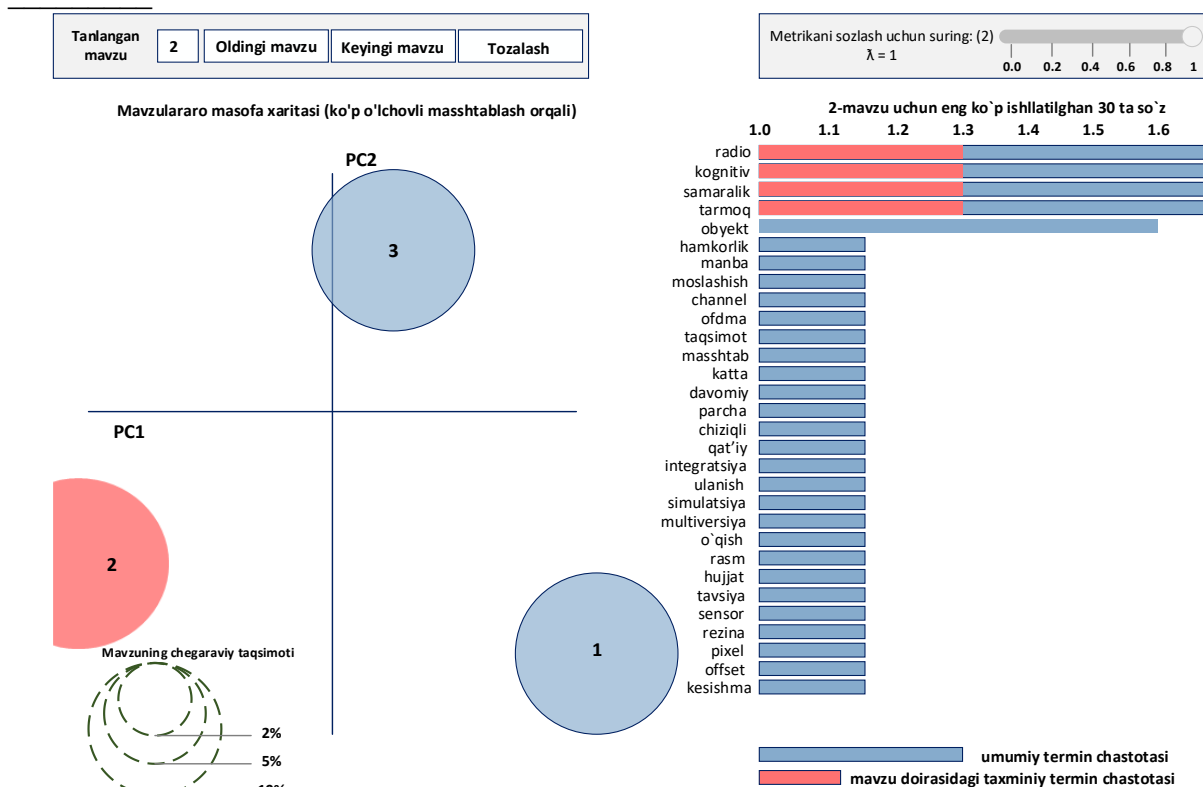
LDA usuli orqali aniqlangan mavzularni vizualizatsiya turli Python paketlarida mavjud bo'lgan turli xil usullar yordamida amalga oshirilishi mumkin. Shuning uchun vizualizatsiya grafigi farqlanishi mumkin, ammo natija bir xil bo'ladi. U mavzularning xilma-xilligi va ma'lumotlarda yoki turli hujjatlarda taqsimlanishi mumkin. Hatto **t-SNE** kabi o'lchamlarni kamaytirish usullari ham turli hujjatlardagi terminlardan foydalangan holda bashorat sifatida ishlatilishi mumkin. Tematik modellash tirishni namoyish qilish uchun ishlatiladigan ba'zi Python paketlari: **sklearn**, **gensim**.



3-rasm. Korpus matn(hujjat)larini tematik modellash tirish

Tematik modellash tirishdagi ma'lumotlar vizualizatsiyasi

```
import pyLDAvis.sklearn
pyLDAvis.enable_notebook()
pyLDAvis.sklearn.prepare(best_lda_model,
small_document_term_matrix,small_count_vectorizer,mds='tsne')
```



4-rasm. Tematik modellash tirishdagi ma'lumotlar vizualizatsiyasi

Xulosa

Tematik modellashtirish – katta hajmdagi til korpusi matnlaridagi mavzularni o‘rganishga qaratilgan nazoratsiz mashinali o‘rganish modellari asosida so‘zlar va hujjatlarning vaznli ro‘yxatini aniqlash usuli. Tematik modellashtirish – sun‘iy intellektdan foydalanadigan tabiiy tilni qayta ishlash texnologiyasi bo‘lib, turli tashkilotlarga ish tizimini boshqarish, mijozlarga uzluksiz xizmat ko‘rsatish va turli jarayonlarda qulay sharoit yaratish uchun xizmat qiladi. Tematik modellashtirish usullari vositasida tashkilotdagi jarayonlarni avtomatlashtirish va mablag‘larni tejash uchun sun‘iy intellekt metodlaridan foydalangan holda tashkilot faoliyati samaradorligini oshirish mumkin. Ushbu maqolada NLPda tematik modellashtirish usulidan foydalangan holda til korpusi matnlaridagi mavzularni aniqlash, natijalarni tahlil qilish va talqin qilish qisqacha tavsiflandi.

Foydalanilgan adabiyotlar:

1. B.Elov, Sh.Khamroeva, Z.Xusainova (2023). The pipeline processing of NLP. *E3S Web of Conferences* 413, 03011, INTERAGROMASH 2023. <https://doi.org/10.1051/e3sconf/202341303011>
2. Elov, B., Xusainova, Z., & Berdiyeva, H. (2023). The Problem of Words Undergoing Sound Changes in Uzbek Stemmers. *Central Asian Journal of Literature, Philosophy and Culture*, 4(6), 107-114. Retrieved from <https://cajlp.centralasianstudies.org/index.php/CAJLPC/article/view/905>
3. B.Elov, Sh.Hamroyeva, X.Axmedova. Methods for creating a morphological analyzer. *14th International Conference on Intelligent Human Computer Interaction, IHCI 2022, 19-23 October 2022, Tashkent*. https://dx.doi.org/10.1007/978-3-031-27199-1_4
4. Elov, B., Hamroyeva, S., Alayev, R., Xusainova, Z., & Yodgorov, U. (2023). O‘ZBEK TILI KORPUSI MATNLARINI QAYTA ISHLASH USULLARI. *DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE*, 1(3), 117–129. Retrieved from <https://dtai.tsue.uz/index.php/dtai/article/view/v1i317>
5. B.Elov, E.Adalı, Sh.Khamroeva, O.Abdullayeva, Z.Xusainova, N.Xudayberganov (2023). The Problem of Pos Tagging and Stemming for Agglutinative Languages. *8 th International Conference on Computer Science and Engineering UBMK 2023, Mehmet Akif Ersoy University, Burdur – Turkey*.
6. Z.Xusainova. O‘zbek tili milliy korpusi qidiruv tizimini optimallashtirishda lemmatizatsiyadan foydalanish. *O‘zbekiston: Til va Madaniyat (Kompyuter lingvistikasi)*, 2023, 2(6). ISSN 2181-922X
7. B.Elov, Sh..Hamroyeva, Z.Xusainova. O‘zbek tilidagi turli tuzilishli so‘zlarni lemmalash usullari. *O‘zbek tili milliy va ta’limiy korpusining nazariy va amaliy masalalari mavzusidagi respublika ilmiy-amaliy konferensiyasi materiallari*, 2023-yil 5-may
8. B.Elov, Z.Xusainova, N.Xudayberganov. O‘zbek tili korpusi matnlari uchun TF-IDF statistik ko‘rsatkichni hisoblash. *SCIENCE AND INNOVATION INTERNATIONAL SCIENTIFIC JOURNAL VOLUME 1 ISSUE 8 UIF-2022: 8.2 | ISSN: 2181-3337*
9. Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11). <https://doi.org/10.1007/s11042-018-6894-4>
10. Kherwa, P., & Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, 7(24). <https://doi.org/10.4108/eai.13-7-2018.159623>

11. Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. *ACM International Conference Proceeding Series*, 382. <https://doi.org/10.1145/1553374.1553515>
12. Korencic, D., Ristov, S., Repar, J., & Snajder, J. (2021). A Topic Coverage Approach to Evaluation of Topic Models. *IEEE Access*, 9. <https://doi.org/10.1109/ACCESS.2021.3109425>

ROBERTA TIL MODELI VA NMF TEMATIK MODELLASHTIRISH USULI

Aloyev Narzillo Raxmatilloevich

Alisher Navoiy nomidagi

Toshkent davlat o'zbek tili va adabiyoti universiteti

tayanch doktorant,

E-mail: vip.alayev@gmail.com

Annotation: Online reviews now provide customers with relevant information for the purchase decision process. Such reviews change the way online consumers make decisions and form their impressions of a product, as well as affect the effectiveness of product sales. According to Consumer Reports, a product without online reviews is purchased 270% less often than a product with 5 positive online reviews. In addition, a public survey shows that 98% of online consumers use online reviews as information when making online purchases and spend more than a minute reading online reviews before making a purchase. However, buyers are often skeptical of reviews that are too numerous. Studies show that 82% of customers specifically look for negative reviews. It can be concluded that reviews, in turn, play a very important role in making a purchase decision. Therefore, the following is an analysis of reviews, their sentiment analysis using various methods, to determine what prompted the customer to make such a review, and to provide product providers with the right solution for each case. Recent studies have worked in the areas of sentiment analysis, comparing the different available methods, and identifying the methods that give the best results, but none of them have made significant progress in identifying the subject matter of reviews left by customers. This is where the uniqueness of this work comes in. This study not only determines whether a review is positive or negative, but also models the subject matter of the reviews. We find answers to the questions of identifying the elements that affect customer satisfaction and how to make the right decision in each case.

Keywords: Roberta, Bert, LSTM, NMF, NLP, language models

Аннотация: сегодня онлайн-обзоры предоставляют покупателям важную информацию для принятия решения о покупке. По данным Consumer Reports, вероятность покупки товара без онлайн-отзывов на 270% ниже, чем у товара с 5 положительными онлайн-отзывами. Кроме того, опрос общественного мнения показывает, что 98% интернет-потребителей используют онлайн-отзывы в качестве информации при совершении покупок в Интернете и тратят более минуты на чтение онлайн-отзывов перед совершением покупки. Однако покупатели часто скептически относятся к слишком многочисленным отзывам. Исследования показывают, что 82% клиентов специально ищут отрицательные отзывы. Можно сделать вывод, что отзывы, в свою очередь, играют очень важную роль в принятии решения о покупке. Поэтому нижеследующая цель — проанализировать отзывы, проанализировать их настроение с помощью различных методов, определить, что побудило клиента написать такой отзыв, и предоставить поставщикам продукции правильное решение для каждой ситуации. Недавние исследования проводились в области анализа настроений, сравнения различных существующих методов и выявления методов, дающих наилучшие результаты, однако ни один из них не достиг существенного прогресса в определении предмета отзывов клиентов. Вот тут-то и проявляется уникальность этой работы. Это исследование не только определяет, является ли отзыв положительным или отрицательным, но и моделирует темы отзывов. Мы найдем ответы на такие вопросы, как определение элементов, влияющих на удовлетворенность клиентов, и как принять правильное решение в каждом конкретном случае.

Ключевые слова: Roberta, Bert, LSTM, NMF, NLP, языковые модели.

Annotatsiya: Hozirgi kunda onlayn sharhlar mijozlarga xarid qilish to'g'risida qaror qabul qilish jarayoni uchun tegishli ma'lumotlarni taqdim etadi. Bunday sharhlar onlayn iste'molchilarning qaror qabul qilish usullarini o'zgartiradi va mahsulot haqidagi taassurotlarini shakllantiradi, shuningdek, mahsulotni sotish samaradorligiga ta'sir qiladi. Consumer Reports ma'lumotlariga ko'ra, onlayn sharhlarsiz mahsulot 5 ta ijobiy onlayn sharhga ega bo'lgan mahsulotga qaraganda 270% kamroq sotib olinadi. Bundan tashqari, ommaviy so'rov shuni ko'rsatadiki, onlayn iste'molchilarning 98 foizi onlayn xaridlarni amalga oshirishda onlayn sharhlardan ma'lumot sifatida foydalaniladi va xarid qilishdan oldin onlayn sharhlarni o'qish uchun bir daqiqadan ko'proq vaqt sarflanadi. Biroq xaridorlar, soni juda ham ko'p bo'lgan sharhlarga ko'pincha shubha bilan qarashadi. Tadqiqotlar shuni ko'rsatadiki, mijozlarning 82% ayniqsa salbiy sharhlarni qidiradi. Xulosa qilish mumkinki, sharhlar o'z navbatida xarid qilish qarorini qabul qilishda juda ham katta rol o'ynaydi. Shu sababli, quyida sharhlarni tahlil qilish, turli usullar yordamida ularning sentiment tahlil qilish, mijozni bunday ko'rib chiqishga nima undaganligini aniqlash va mahsulot yetqazib beruvchilarga har bir holat uchun to'g'ri yechimni taqdim etishdir. So'nggi tadqiqotlar hissiyotlarni tahlil qilish, turli xil mavjud usullarni taqqoslash va eng yaxshi natijalarni beradigan usullarni aniqlash sohalarida ishladi, ammo ularning hech biri mijozlar qoldirgan sharhlar mavzusini aniqlash uchun oldinga sezilarli darajada siljimadi. Bu ishning o'ziga xosligi shu yerda paydo bo'ladi. Ushbu tadqiqot sharhning ijobiy yoki salbiy ekanligini aniqlabgina qolmay, sharh mavzularini modellashtiradi. Mijozlarning qoniqishiga ta'sir qiluvchi elementlarni aniqlash va har bir holatda to'g'ri qaror qanday qabul qilinishi kerak degan savollarga javob topamiz. Sabablarni tahlil qilish o'z navbatida mijozlarning qoniqishini oshirish va ehtimol noroziliklar sonini kamaytirish to'g'risida kerakli qaror qabul qilishga yordam beradi.

Kalit so'zlar: Roberta, Bert, LSTM, NMF, NLP, til modellari

Maqola quyidagicha tuzilmaga ega: birinchi bo'lib qo'llaniladigan usulning umumiy tushunchasini keltirilgan va foydalanilgan tushunchalarning ta'riflari to'plangan. Shundan so'ng ikki qismga bo'lingan turli xil usullarning amalga oshirish qismi keltiriladi. Birinchi, *LSTM*, *GRU*, *Bi-LSTM* va *Roberta* orasida eng yaxshi modelni tanlash hamda, mijozlarning fikr-mulohazalarini tahlil qilish. Ikkinchisi esa tematik modellashtirishga tegishli. Nihoyat, natijalarni vizualizatsiya qilish va to'g'ri yechimlarni taklif qilish.

NLP - bu kompyuterlar tabiiy tildagi matnni qanday tushunishini o'rganadigan tadqiqot sohasi. Katta hajmdagi matn ma'lumotlarini odamlar qayta ishlay olmasligi sababli, NLPning asosiy maqsadi inson aqlini o'qish va talqin qilish orqali inson tillarini tezroq va osonroq qayta ishlashdir. Kompyuterlarga tabiiy tillarni tushunish imkonini berish orqali kerakli maqsadlarga erishish uchun hozirda ko'plab vositalar va texnikalar ishlab chiqilgan. Ushbu ishda matn xarakteristikasini ajratib olishning ikki xil arxitekturasini bilan tajriba o'tqaziladi: neyron tarmoqlar va PLM (oldindan o'rgatilgan til modellari – Pre-trained language models).

RNN(Recurrent neural networks – Rekurent neyron tarmoqlar)

RNN ma'lumotlarning qisqa ketma-ketligini o'rganishda juda yaxshi bo'lib hisoblanadi. Ammo yo'qolib borayotgan gradiyent muammosi ularning uzoq va murakkab kontekstlarni o'rganish va tushunish qobiliyatiga to'sqinlik qiladi. Ushbu muammoni hal qilish uchun RNN ning maxsus turi - "Uzoq qisqa muddatli xotira" ishlab chiqilgan ya'ni, LSTM.

Uzoq qisqa muddatli xotira (LSTM): **LSTM** - bu yashirin qatlamlardagi yacheykalarga ega bo'lgan nazorat qilinadigan takrorlanuvchi tarmoq bo'lib, dastlab matnning hissiyotlarini tasniflash uchun ishlatilgan va tabiiy tilni qayta ishlashda keng qo'llanilgan.

Gated Recurrent Unit (GRU): **GRU** bu o'z navbatida LSTM ning qisqartirilgan versiyasi bo'lib, bunda tarmoq mahsuldorligi yaxshilanadi va o'qitish vaqti qisqaradi. GRU va LSTM bir xil uslubda ishlaydi, lekin ular "gate" chastotasi va yacheyka holatini saqlashda farqlanadi.

Ikki tomonlama uzoq qisqa muddatli xotira (Bi-LSTM): **Bi-LSTM** joriy etilishi bilan oq sentiment tahlili yaxshilandi. Bi-LSTM - bu LSTM ning yanada barqaror versiyasi bo'lib hisoblanadi.

PLM(Pre-trained language models – oldindan o‘qitilgan til modellari)

So‘nggi yillarda NLP bo‘yicha Deep Learning ommaviy tus oldi. Deep Learning modellari o‘zlarini NLP vazifalarida yaxshi natijalarga ega ekanligini isbotlagan bo‘lsa-da, ularning juda ham katta ma‘lumotlar to‘plami talab qilinishi tanqidlarga yo‘l ochib berdi. Ular o‘rgatilishi uchun katta hajmdagi ma‘lumotlarni talab qiladi. Ushbu muammoni hal qilish uchun “oldindan o‘qitish” deb nomlanuvchi usul taklif qilindi.

RoBERTa: So‘nggi paytlarda oldindan o‘qitilgan modellarning (PTM) o‘sishi tabiiy tilni qayta ishlashda (NLP) yangi davrni belgiladi. Facebook tomonidan taklif qilingan RoBERTa bu - BERT modelining takomillashtirilgan kengaytmasi hisoblanadi. BERT modeli mavzuni modellashtirish va hissiyotlarni tahlil qilish kabi ko‘plab sohalarda qo‘llaniladi. Facebook AI Research (FAIR) uni nusxalash va giperparametrlarning uning ishlashiga ta‘sirini tahlil qilganidan so‘ng, uni kam o‘qitilgan deb aniqladi. Shunday qilib, BERTning takomillashtirilgan va kuchli versiyasi bo‘lgan RoBERTa yaratildi.

Tematik Modellashtirish

Mavzu modellari - bu ma‘lumotlarda mavjud bo‘lgan turli mavzularni aniqlaydigan va foydalanuvchiga har bir hujjatda yoritilgan mavzularni tushunishga yordam beradigan “unsupervised machine learning”ning tasdiqlangan usullaridir. Mavzuni modellashtirish tabiiy tilni qayta ishlash-NLP va axborot qidirish-IR kabi turli sohalarda qo‘llaniladi.

NMF - matn matritsasi atamaları ikkita matritsaga: hujjat-mavzu va atama mavzusiga bo‘linadigan usuldir. W belgisi hujjatning faktor matritsasi tasviri, H belgisi esa atamaning faktor matritsasi ko‘rinishidir. NMF mavzularni modellashtirishning mashhur usuli bo‘lib, matnlar orasidagi yashirin aloqalarni topish va yoritilgan mavzularni aniqlash uchun keng qo‘llaniladi. Bundan tashqari, u ma‘lumotlarni taqdim etish va mavzuni modellashtirishda juda samarali deb tan olingan.

Ma‘lumotlarni tayyorlash

Ma‘lumotlarni oldindan qayta ishlash modelni o‘qitishda muhim qadam bo‘lib hisoblanadi, chunki u modelning ishlashida muhim rol o‘ynaydi. Matnni har qanday mashinani o‘rganish algoritimga berishdan oldin uni tozalash kerak. Norozi yoki jahli chiqqan mijozlar sharhlarida mavjud bo‘lgan “shovqin”, matnning semantikasini tushunishni va foydali ma‘lumotlarni olishni qiyinlashtiradi. Shovqinlar emotsiya ko‘rinishidagi “emoji”lar yoki qisqa jargonlar bo‘lishi mumkin. Ma‘lumotlar tozalangan so‘ng, matn vektor ko‘rinishiga olib kelinadi. Quyidagi oldindan qayta-ishlangan ma‘lumotlar ishlatiladi:

- Kichik harflardan iborat matn: Katta va kichik harflar mashina tomonidan turlicha talqin qilinadi. Shunday qilib, barcha so‘zlarni kichik harflarga aylantirish katta/kichik harflar muammosini hal qilish uchun muhimdir.
- “Probellarni” normalizatsiya qilish: parallel bo‘shliqlarni (probellarni) bitta bo‘shliq bilan almashtirish
- Faqat harflarni qoldirish: “emojilar”, tabassumlar va tinish belgilari olib tashlanadi, chunki mashina ularni ko‘pincha so‘z sifatida qabul qiladi.
- Takroriy belgilarni olib tashlash: mijozlar o‘z fikrlarini bildirish uchun belgilarni "haqiqattan ham yomon" kabi so‘zlar bilan takrorlaydilar. Ushbu bosqichda takroriy belgilar o‘chiriladi.
- Matnni tokenlarga ajratish: Tokenizatsiya PLNni oldindan qayta ishlashda keng tarqalgan qadamdir. Matndagi har bir so‘zni (tokens) ajratish va ularni ro‘yxatda saqlash uchun ishlatiladi. Ikki nusxali tokenlarni olib tashlash va unikallarni yaratish o‘z navbatida ma‘lumotlar miqdorini kamaytiradi.
- “STOP” so‘zlari bilan tokenlarni filtrlash: STOP so‘zlari ko‘pincha ustunlik, yuklama va fe‘llarni o‘z ichiga oladi. Bu so‘zlar matnni tasniflash uchun hech qanday foydali ma‘lumotni taqdim etmaydi. Aksincha, bu so‘zlar modelning ishlashiga ta‘sir qilishi mumkin bo‘lgan “shovqin” hosil qiladi. NLTK kutubxonasi matndan keraksiz

ma'lumotlarni olib tashlash uchun havola sifatida ishlatilishi mumkin bo'lgan bo'sh so'zlar ro'yxatini taqdim etadi.

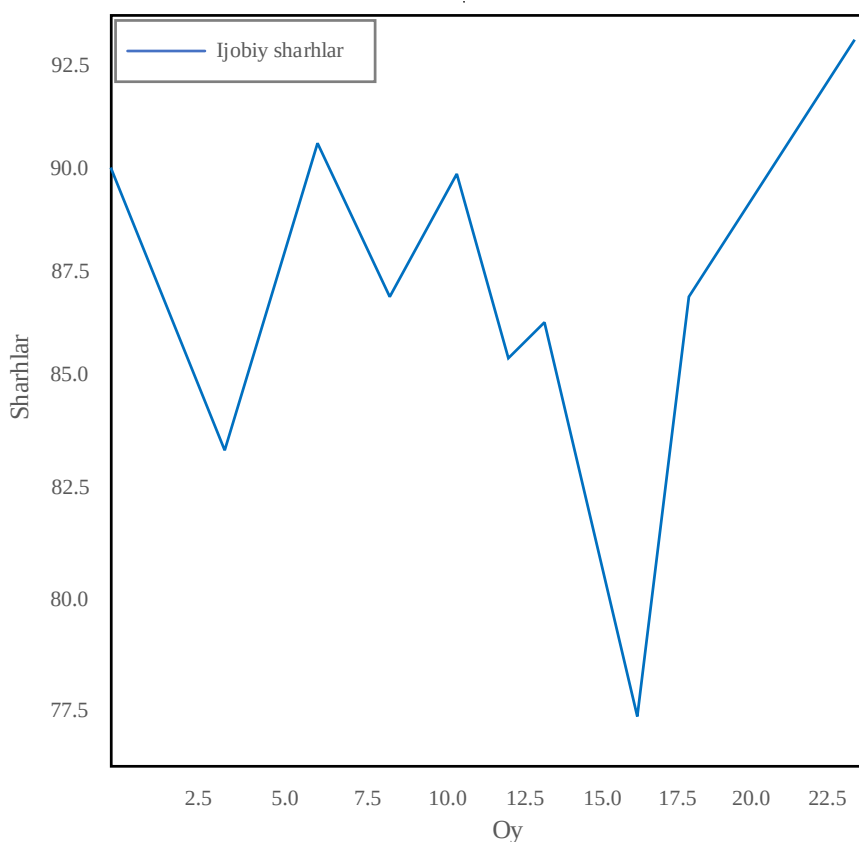
- Lemmatizatsiya texnikasini so'zlarga qo'llash: u turli shakllardagi bir xil so'zlarni asosiy shaklga aylantirishdan iborat. Lemmatizatsiya kontekstni hisobga oladi va bir xil so'zning turli shakllarini bitta mazmunli asos shakliga aylantiradi.
- Matnning vektor ko'rinishini qo'llash: so'zlar ketma-ketligi bo'lgan asl matn ko'rinishi mashinani o'rganish algoritmlari uchun tushunarli emas. Bu algoritmlar sonli vektor tasvirlarini olishlari kerak. Bu yerda vektor tasvirining roli asosiy vazifani bajaradi, har bir matnni raqamlar ketma-ketligiga aylantiradi. Bu muammoni hal qilishning TF-IDF mashhur va eng oddiy usuldir.

Hujjat chastotasining teskari atama chastotasi (TF-IDF): Bu matnli ma'lumotlardan ma'lumot olish uchun moliyaviy sohadagi ajoyib ilovalarga ega statistik yondashuv. Bu usul odatda tabiiy tilni qayta ishlash, axborot qidirish va matnni ajratib olishda qo'llaniladi. Ma'lumot olish va ko'plab matn ajratish vazifalari uchun TF-IDF samarali muddatli ajratib olish sxemasi sifatida tan olingan. U hujjatlar to'plamidagi atamalarning ahamiyatini baholovchi vaznni va o'lchovni belgilaydi. Matndagi so'zning TF-IDF ko'rsatkichi 1-da ko'rsatilganidek hisoblanadi.

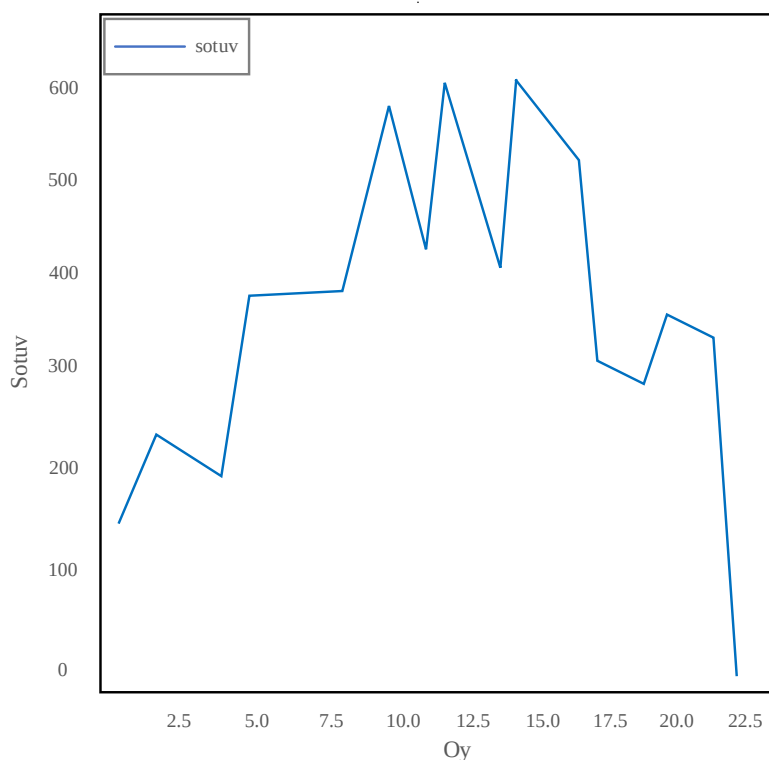
$$W_{x,y} = (tf_{x,y}) \times \left(\log\left(\frac{N}{df_x}\right) \right) \quad (1)$$

Atama chastotasi x atamasi y ($tf_{x,y}$) hujjatida qanchalik tez paydo bo'lishini ko'rsatadi.

Hujjatning teskari chastotasi barcha ($\log(N/df_x)$ hujjatlar kontekstida x atamasining ahamiyatini o'lchaydi, df_x - x mavjud bo'lgan hujjatlar soni va N - hujjatlarning umumiy soni). Boshqacha qilib aytganda, TF-IDF uchun matnning ahamiyati hujjatdagi hodisalar soniga mutanosib ravishda ortadi.



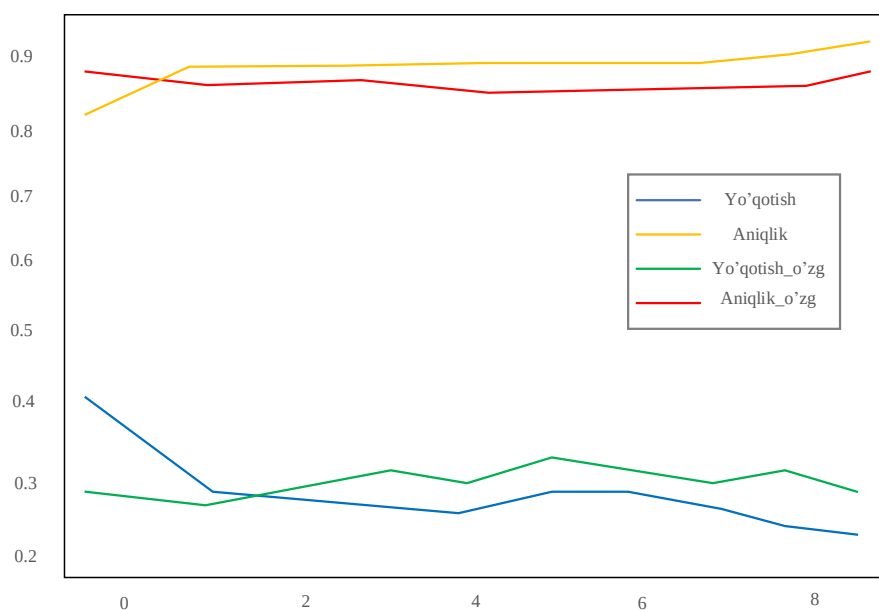
1-Rasm. Oylik ijobiy sharhlar



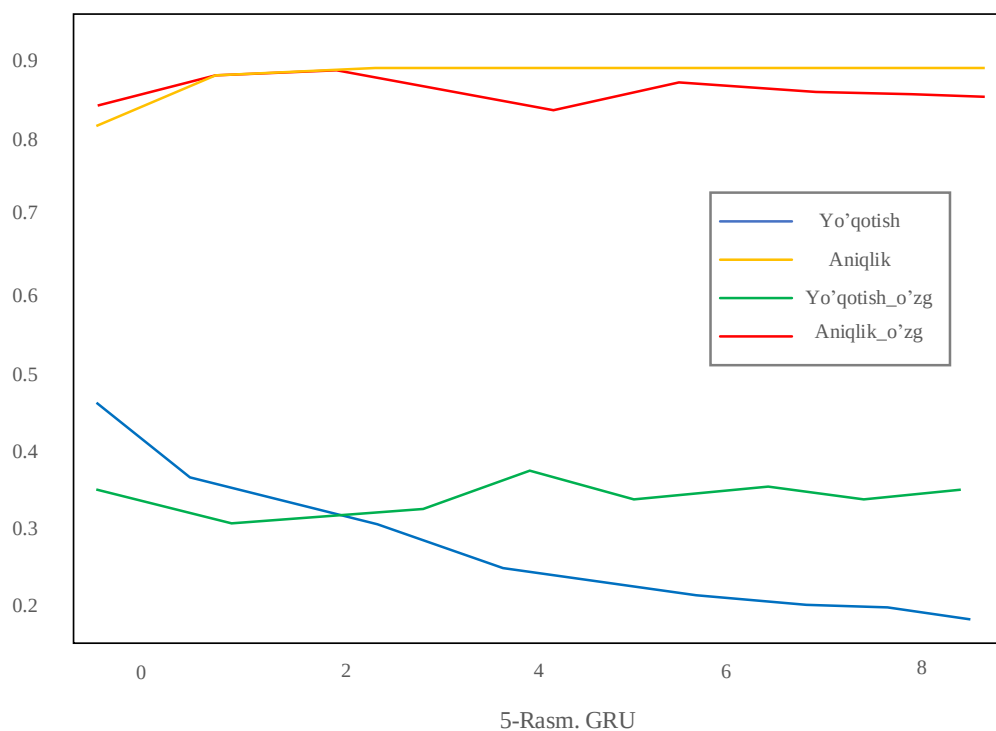
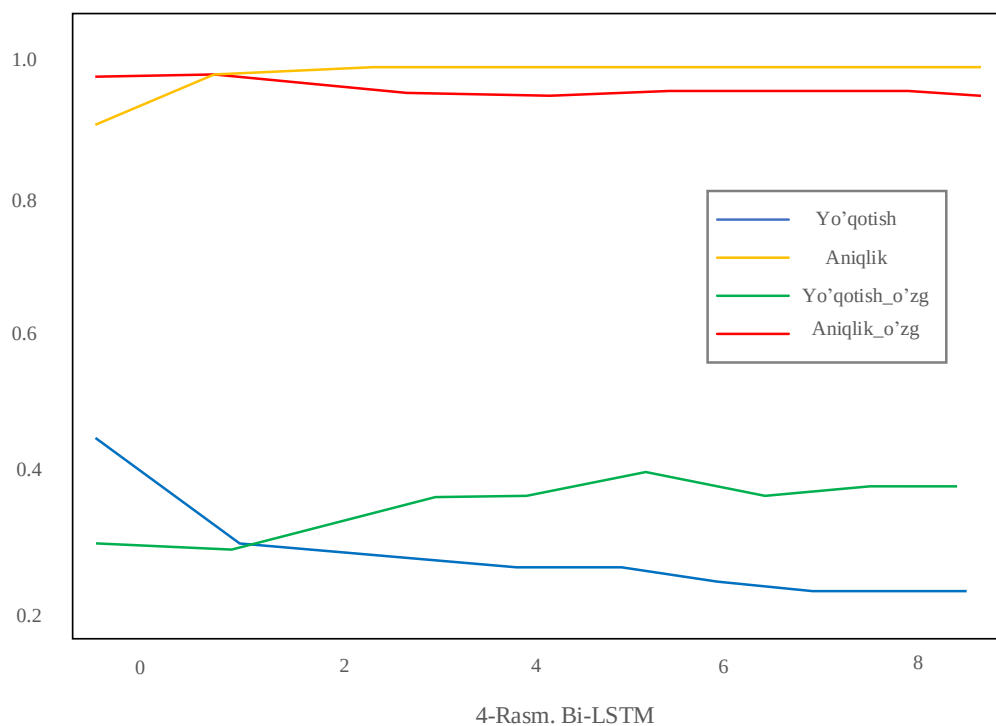
2-Rasm. Oylik sotuvlar

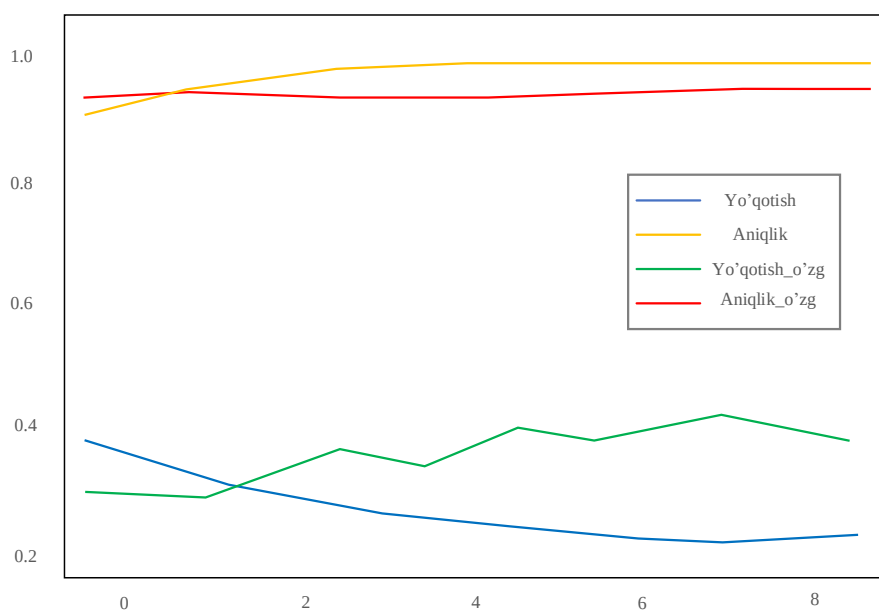
Natijalar vizualizatsiyasi va tahlili

TF-IDF matritsasi turli xil tabiiy tillarni qayta ishlash modellarining dastlabki ma'lumotlarni kiritish jarayoni hisoblanadi. LSTM, GRU, BiLSTM va Roberta modellaridan mos algoritmlar tanlanadi. 3, 4, 5 va 6-rasmlarda o'rganish chiziqlari ko'rsatilgan. Barcha modellar (LSTM, Bi-LSTM, GRU va Roberta) uchun o'quv to'plamining aniqligi yo'qotishning doimiy pasayishi bilan birga o'sishda davom etmoqda. Sinovning aniqligi Roberta modeli bilan maksimal darajaga yetadi.



3-Rasm. LSTM





6-Rasm. RoBERTa

Mahsuldorlik metrikalari

Turli modellar 1-jadvalda ko'rsatilganidek, to'rtta baholash ko'rsatkichlari yordamida baholanadi. Ushbu jadval aniqlik, eslab qolish, F1 va aniqlikning makro-o'rtacha qiymatlarini o'z ichiga oladi. Kutilganidek, PLM (Roberta) modeli ko'pgina ko'rsatkichlar bo'yicha RNN modellariga qaraganda yaxshiroq ishladi. RNN modellari orasida Bi-LSTM o'z navbatida LSTM va GRUga qaraganda biroz yaxshiroq ish ko'rsatdi.

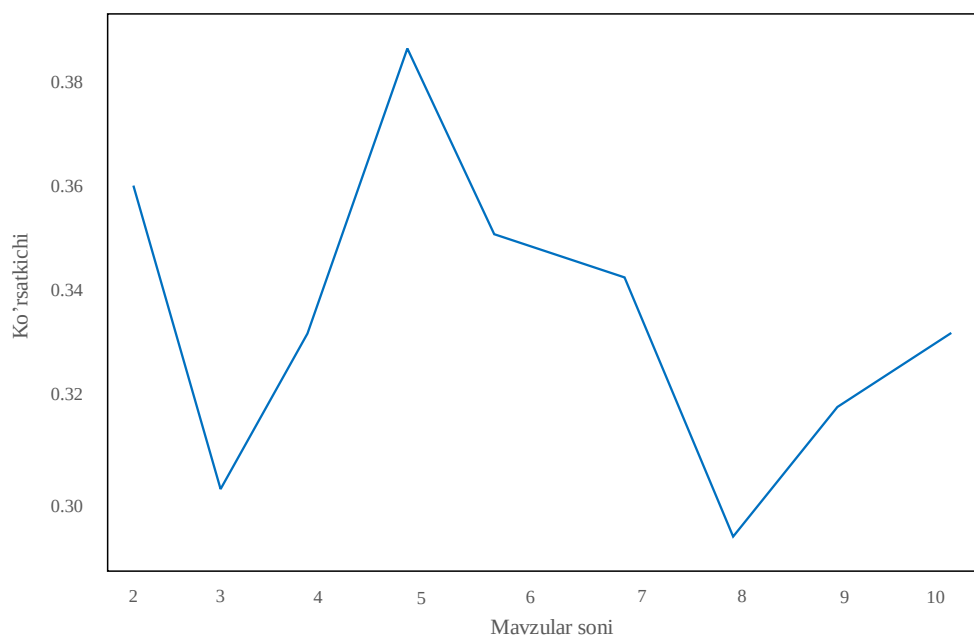
1-jadval LTSM,GRU,Bi-LTSM va RoBERTa
natijalarining ko'rsatkichlari

Model	Aniqligi	Xotira	F1	Ko'rsatkichi
<i>LSTM</i>	0.95	0.91	0.93	0.95
<i>GRU</i>	0.97	0.92	0.94	0.96
<i>Bi-LSTM</i>	0.98	0.95	0.96	0.97
<i>RoBERTa</i>	0.98	0.97	0.97	0.98

Tematik modellashtirish natijalari

Sharhlarning xususiyatini va ularning ijobiy va salbiy foizini aniqlagandan so'ng, biz tematik modellashtirish bosqichiga o'tamiz. Bu bosqichda birinchi navbatda 2 dan 10 gacha bo'lgan oraliqdagi mavzularning eng yaxshi soni aniqlanadi, so'ngra ko'rib chiqish matnining

tematik tuzilishi aniqlanadi. Mavzularning ixtiyoriy sonini topish uchun muvofiqlik ko'rsatkichi hisoblab chiqiladi.



7-Rasm. NMF mavzular soniga asosan ko'rsatkich o'zgaruvchilari

7-rasmda natijalarning vizualizatsiyasi ko'rsatilgan. Shunga asoslanib, NMFga kiritish uchun optimal raqam 5 hisoblanadi, chunki bu raqam eng yuqori muvofiqlik ko'rsatkichiga ega. Mavzular soni aniqlangandan so'ng, 3-jadvalda ko'rsatilganidek, beshta alohida, yashirin so'z klasterlarini aniqlash uchun mavzularni modellashtirish uchun NMF algoritmi qo'llaniladi. Har bir klaster alohida mavzuga tegishli. Har bir klaster nomini aniqlash uchun 2-jadvalda ko'rsatilganidek, har bir klasterda eng ko'p qo'llaniladigan so'zlar aniqlanadi va tahlil qilinadi.

2-jadval. Tematik modellashtirish uchun
ma'lumotlar to'plami namunasi

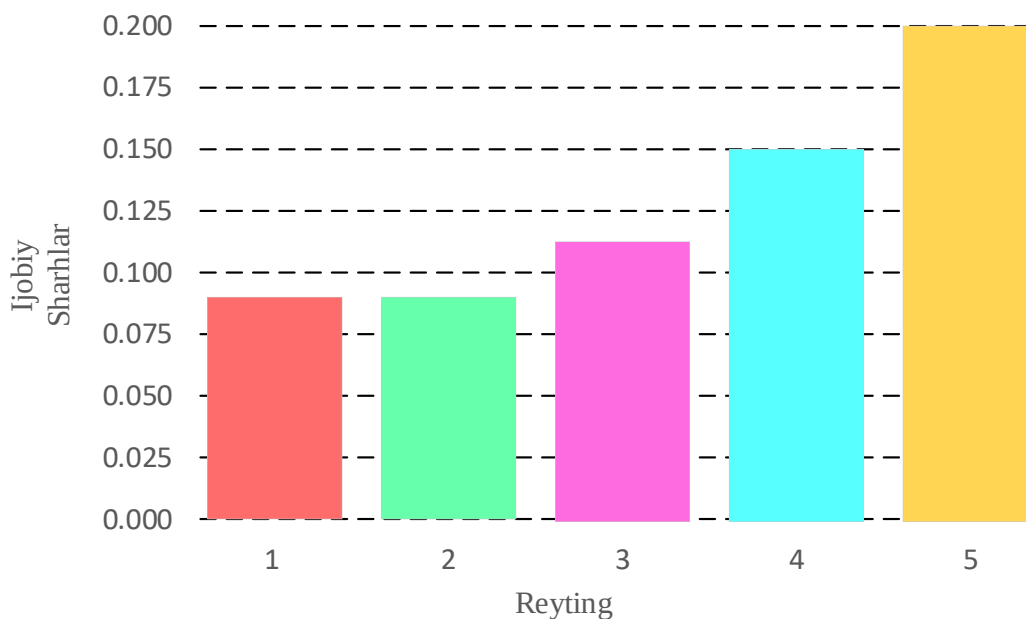
Sharhlar ID	Kontent	Sentiment tahlil	%	%	Mavzu ID
2...8	Men uni kelishilgan sanadan ancha oldin oldim	ijobiy	0.1	99.9	0
7...e	Sifati judayam zo'r mahsulot	ijobiy	0	100	1
1...c	Men pulni qaytarishni so'radim lekin, hech kim indamadi	salbiy	99.9	0.1	2
F...e	Narxi nisbatan to'g'ri kelmaydi lekin, bu yaxshi	ijobiy	37.6	62.4	3
5...d	Men 3 ta xarid qildim lekin, 2 ta oldim	salbiy	99.8	0.2	4

3-jadval. Har bir mavzudagi yuqori chastotaga ega so'zlar

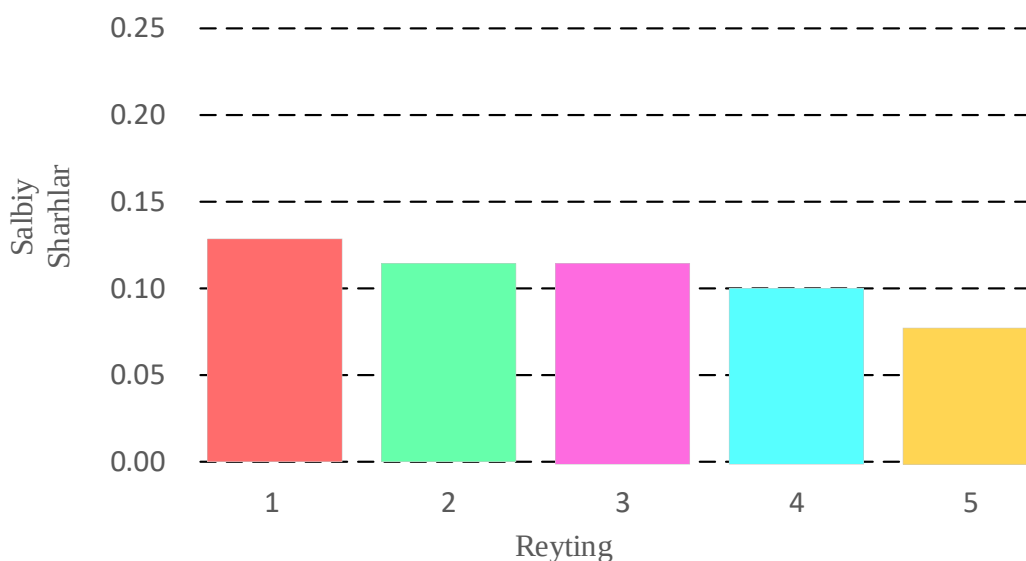
Mavzu ID	Mavzu sarlavhasi	Ko'p uchraydigan so'zlar
0	Vaqt	Vaqt, yetqazish, muddat, yetib borish, olish, tezlik, kutish, kun, yuklash
1	Sifat	Sifat, zarar, ko'rkam, sinish, mo'jaz, arzon, yo'qolish
2	Mijozlarni qo'llab quvvatlash xizmati	Javob, pulni qaytarish, ma'lumot, yordam, bekor qilish, onlayn
3	Narx	Narx, tan narxi, pulni yechib olish, arzon, pulizga arziydi, to'lov
4	Funksionallik	Ish, shovqin, qabul qilib olish, bo'lak, soni, rangi, hajmi

- "Vaqt" mavzusi va yetkazib berish bilan bog'liq tushunchalar mijozlar sharhlarida eng ko'p uchraydigan mavzudir. Barcha sharhlarning 32% ushbu mavzuga bag'ishlangan bo'lib, bu yetkazib berish vaqtlari mavzusining ahamiyatini ko'rsatib beradi.
- Mijozlarning sharhlaridagi ikkinchi mavzu "mijozlarga xizmat ko'rsatish" bilan bog'liq. "Mijozlarga xizmat ko'rsatish" ko'plab mijozlar tomonidan tanqid qilindi, bu barcha sharhlarning 26% ni tashkil qiladi.
- Mijozlarning sharhlari orasida uchinchi asosiy mavzu - bu "sifat". Chastotasi bo'yicha "sifat" mavzusi umumiy sharhlar sonining 20% ni tashkil qiladi.
- To'rtinchi mavzu mahsulotning "funksionalligi" bilan bog'liq. Sharhlarning 13% i to'g'ridan-to'g'ri ushbu mavzuni yoritib beradi.
- Va nihoyat, mijozlar sharhlaridagi beshinchi mavzu bu - "narx". Ushbu mavzu barcha sharhlarning 9% ini tashkil qiladi.

Yomon sharh har doim ham yomon baho bilan birgalikda "yurmaydi" va aksincha. Bu, asosan, mijozning shaxsiyati va uning baholashga bo'lgan nuqtai nafari bilan bog'liq. 5 shkaladan iborat baholashda 3 qiymati ba'zilar uchun yaxshi, boshqalar uchun yomon deb hisoblanadi. Bundan tashqari, yaxshi mijozlar ko'pchilik yomon sharh qoldirgan bo'lsa ham yaxshi baho berishadi. 8-rasmda va 9-rasmda ko'rinib turibdiki, reyting oshgani sayin ijobiy sharhlar ko'payadi va reyting oshgani sayin salbiy sharhlar kamayadi.



8-rasm. Ijobiy sharhlar va reyting muqobilligi



9-rasm. Salbiy sharhlar va reyting muqobilligi

Xulosa qilib aytish mumkinki, ushbu maqolada bajarilgan ishlar ikki bosqichdan iborat. Birinchidan, mijozlar qoldirgan sharhlarning xususiyatlarini aniqlash uchun semantik tahlil qilish amalga oshiriladi. Shu maqsadda bir nechta modellardan foydalanilgan va baholangan. Ushbu modellar orasida RoBERTa eng yaxshi ko'rsatkichni ko'rsatdi. Ikkinchi bosqich modellashtirish bosqichi fikrlarning xususiyatlarini va ularning ijobiy va salbiy foizini aniqlagandan so'ng boshlanadi. Ushbu bosqich matnning tematik tuzilishi ya'ni, mijozlar ehtiyojini qondirishga ta'sir qiluvchi elementlarni aniqlashdan iborat. Umuman olganda, sharhlar soni xaridlarning umumiy soniga nisbatan past nisbatdadir, chunki barcha mijozlar o'z fikrlarini baham ko'rishga odatlanmagan.

Foydalanilgan adabiyotlar:

1. Elov B., Alayev R., Aloyev N. (2024). Tematik modellashtirishning zamonaviy usullari. *Digital transformation and artificial intelligence*, 2(1), 8–16. Retrieved from <https://dtai.tsue.uz/index.php/dtai/article/view/v2i12>
2. Elov.B., Alayev N. Matnlarini tematik modellashtirish va tasniflash usullari. barqarorlik va yetakchi tadqiqotlar onlayn ilmiy-amaliy jurnali. Vol. 3 No. 12 (2023). 263-276
3. Elov B., Aloyev N., Yuldashev A. (2023). SVD va NMF metodlari orqali tematik modellashtirish. O`zbekiston: til va madaniyat (Kompyuter lingvistikasi), 2023, 2(6). 55-66
4. Alghamdi, R., & Alfalqi, K. (2015). A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications*, 6(1). <https://doi.org/10.14569/ijacsa.2015.060121>
5. Tao, R., Wei, Y., & Yang, T. (2021). Metaphor Analysis Method Based on Latent Semantic Analysis. *Journal of Donghua University (English Edition)*, 38(1). <https://doi.org/10.19884/j.1672-5220.202010087>
6. Darmalaksana, W., Slamet, C., Zulfikar, W. B., Fadillah, I. F., Maylawati, D. S. adillah, & Ali, H. (2020). Latent semantic analysis and cosine similarity for hadith search engine. *Telkomnika (Telecommunication Computing Electronics and Control)*, 18(1). <https://doi.org/10.12928/TELKOMNIKA.V18I1.14874>
7. Ke, Z. T., & Wang, M. (2022). Using SVD for Topic Modeling. *Journal of the American Statistical Association*.
8. <https://doi.org/10.1080/01621459.2022.2123813>
9. Churchill, R., & Singh, L. (2022). The Evolution of Topic Modeling. *ACM Computing Surveys*, 54(10). <https://doi.org/10.1145/3507900>
10. Kherwa, P., & Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, 7(24). <https://doi.org/10.4108/eai.13-7-2018.159623>

KORPUS UCHUN LUG‘AT TO‘PLASH VA ULARNI TASNIFLASH

Botirova Maftuna

O‘zbekiston Milliy universiteti 4-bosqich talabasi

E-mail: mbotirova501@mail.com

Qodirova Zebo Gulboyevna

O‘zbekiston Milliy universiteti o‘qituvchisi

E-mail: qodirovazebo5450@gmail.com

Annotatsiya: Maqolada ta’limiy korpuslarning lingvistik tadqiqotlar va til o‘rgatish jarayonidagi o‘rni tahlil qilinadi. Xususan, ingliz, yunon, turk va o‘zbek tillariga oid yirik ta’limiy korpuslar jumladan, International Corpus of Learner English, British National Corpus, Cambridge International Corpus, Turkish Student Corpus va TELCning tarkibi, qo‘llanilish sohasi hamda xatolarni aniqlashdagi roli yoritiladi. O‘zbek tilida yaratilgan ta’limiy korpus misolida matn tanlash, izohli lug‘at tuzish va materiallarni tizimlashtirish bosqichlari bayon etilib, ularning ilmiy va amaliy ahamiyati asoslab beriladi. Tadqiqotda korpuslar orqali o‘quvchilarning til o‘zlashtirish darajasini aniqlash, grammatik xatolarni tahlil qilish va samarali o‘quv materiallarini ishlab chiqish imkoniyatlari namoyon etiladi.

Kalit so‘zlar: Ta’limiy korpus, lingvistik tadqiqot, elektron matnlar, pedagogik metodika, korpus tahlili

Annotation: This article analyzes the role of learner corpora in linguistic research and language teaching. It highlights major educational corpora in English, Greek, Turkish, and Uzbek including the International Corpus of Learner English, British National Corpus, Cambridge International Corpus, Turkish Student Corpus, and TELC focusing on their structure, application areas, and their role in error analysis. The process of creating an educational corpus in Uzbek is discussed, with attention to text selection, compiling a glossary, and systematizing materials, emphasizing both scientific and practical significance. The study demonstrates how corpora can be used to assess learners’ language proficiency, analyze grammatical errors, and develop effective teaching materials.

Keywords: Educational corpus, linguistic research, electronic texts, pedagogical methodology, corpus analysis

Аннотация: В статье рассматривается роль учебных корпусов в лингвистических исследованиях и процессе преподавания языков. Особое внимание уделяется крупным образовательным корпусам на английском, греческом, турецком и узбекском языках таким как International Corpus of Learner English, British National Corpus, Cambridge International Corpus, Turkish Student Corpus и TELC. Анализируется их структура, области применения и значение в выявлении ошибок. На примере учебного корпуса узбекского языка описываются этапы отбора текстов, составления глоссария и систематизации материалов, подчеркивается их научная и практическая значимость. В исследовании показано, как с помощью корпусов можно определить уровень владения языком, анализировать грамматические ошибки и разрабатывать эффективные учебные материалы.

Ключевые слова: Образовательный корпус, лингвистическое исследование, электронные тексты, педагогическая методика, анализ корпусов.

Dastlabki ta’limiy korpusga oid konfrensiya 1994-yilda Lankasterda bo‘lib o‘tgan. Sylviane Granger jamoasi bilan birgalikda ilk bor International copus of learner English ta’limiy korpusi yaratildi. Bu korpusni yaratishdan asosiy maqsad til o‘rganuvchilarning qilgan xaatolarini

o'rganish orqali ularni tuzatish, metodlarini yaratishga qaratilgan edi [Abduraxmonova N, 2023: 56].

Bugungi kunda, dunyoda turli tillar uchun yuzlab yirik korpuslar yaratilgan. Ularning eng mashhurlari qatoriga Longman Spoken and Written English Corpus mualliflar ingliz tilining o'zaki va yozma korpusida 40 milliondan ortiq so'zlarni tahlil qilishga muvaffaq bo'lishdi, matnda u yoki bu grammatik hodisani tasvirlovchi 3000dan ortiq misollar keltirilgan.

The British National Corpus (Yozma va og'zaki ingliz tilidagi so'z chostatalari lug'ati yaratilgan, korpus hajmi 100 million ko'proq so'zdan foydalaniladi), Cambridge International Corpus 1 milliarddan ortiq so'zni o'z ichiga olgan, Korpus bo'yicha og'zaki va yozma ingliz tili grammatikasi bo'yicha "Cambridge grammatikasi" qo'llanmasi ishlab chiqarilgan. Lingvistik korpus asosida yaratilgan lug'atlar orasida Oksford lug'atini olsak, u ingliz korpusiga asoslangan tadqiqotlardagi minglab yangi so'z ma'nolarini aniqlagan [Холиёров Ў, 2021: 14-15].

Hellenic National Corpus zamonaviy yunon tili uchun eng katta korpus 1994-yilda asos solingan bo'lib, 1-versiyasi 1997-yilda 13 million so'zni o'z ichiga olgan. Uning joriy versiyasi 47 million so'zni o'z ichiga oladi [Sidiropulos N, Hatzigeorgiu N, 2011: 4-5].

Ingliz tilida ta'limiy korpus alohida korpus bo'lib, Gonkong Study Corpus ingliz tilini o'rganuvchilarning yozma ishlariga alohida e'tibor qaratadi. Korpus til o'rganuvchilarning yozish jarayonidagi xatolarini, uslublarini va til qoidalariga rioya qilish darajasini o'rganish imkonini beradi. Bu esa ta'lim jarayonida samarali til o'rgatish usullarini ishlab chiqishga xizmat qiladi.

Til o'rgatishga qaratilgan boshqa muhim ta'limiy korpuslar qatoriga Michigan Corpus ham kiradi. U Michigan Universitetida yaratilgan bo'lib, rasmiy va akademik nutq namunalarini o'z ichiga olgan Michigan Corpus of Academic Spoken English (MICASE) nomi bilan mashhur. Ushbu korpus talabalar, professorlar va ilmiy muhitdagi suhbatlarni tahlil qilish orqali akademik nutq xususiyatlarini o'rganishga yordam beradi. Shuningdek, Tampere Universitetida ishlab chiqilgan Tamper Corpus ham muhim manba hisoblanadi. Bu korpus ELFA (English as a Lingua Franca in Academic Settings) – ya'ni akademik muhitda Lingua Franca sifatida ishlatiladigan ingliz tiliga oid real misollarni jamlagan. Korpus turli millatlarga mansub insonlarning akademik sohadagi ingliz tilidan foydalanish uslubini tadqiq qilishga imkon beradi. Mazkur elektron lingvistik resurslarni ishlab chiqishda taniqli tilshunos Anna Mauranen muhim rol o'ynagan. U akademik ingliz tilining o'ziga xos jihatlarni, shuningdek, ona tilidan farqli bo'lgan til tizimlarini o'rganishda ilg'or metodlarni qo'llab keladi. Uning ishlari ta'limiy korpuslar yordamida til o'rgatish va tilshunoslik tadqiqotlarini yanada rivojlantirishga xizmat qiladi [Mauranen Anna and Elina Ranta, 2009: 562].

Turkish Student Corpus (TSC) turk tilini o'rganuvchi talabalarining yozma ishlaridan iborat bo'lgan o'quv korpusidir. Bu korpus turk talabalarining ingliz tilida yozgan 341 ta munozaraviy esselarini o'z ichiga oladi. Yozuvchilarning barchasi ingliz tili o'qituvchiligi bo'yicha ta'lim olayotgan talabalar bo'lib, ularning yozma ishlari o'rta-oliy darajadagi til bilimiga ega [Erdem Akbas, 2022: 4]. Korpusning asosiy maqsadi – ingliz tilini o'rganuvchi turk talabalarining yozma ishlarini tahlil qilish orqali ularning til o'rganish jarayonidagi qiyinchiliklarini aniqlash va bu orqali o'quv materiallarini takomillashtirishdir. Bu kabi o'quv korpuslari til o'rgatishdagi muammolarni tushunish va ularni hal etishda muhim ahamiyatga ega. Shuningdek, turk talabalarining ingliz tilidagi yozma ishlaridagi fe'l xatolarini o'rganish bo'yicha tadqiqotlar ham amalga oshirilgan. Bu tadqiqotlar orqali talabalarining til o'rganish jarayonidagi umumiy xatolari va ularning sabablari aniqlanib, o'quv jarayonini yaxshilashga qaratilgan takliflar ishlab chiqilgan [Cem Can, 2017: 13].

Turk tili-ingliz tili o'rganuvchilar korpusi (Turkish-English Learner Corpus/TELC) – bu ingliz tilini ikkinchi til sifatida o'rganuvchi turk talabalarining yozma matnlaridan iborat korpus bo'lib, uning asosiy maqsadi o'quvchilarning leksik xatolarini aniqlash va tahlil qilishdan iborat. Korpus 231 nafar universitet talabasi tomonidan yozilgan 369 ta matndan iborat bo'lib, umumiy so'zlar soni 104,864 tani tashkil etadi. Ushbu matnlarda 3000 dan ortiq xatolar maxsus belgilangan

va tuzatilgan. TELC ayniqsa turk tilida soʻzlashuvchilar uchun ingliz tilini oʻrganish jarayonidagi xatolarni aniqlash va ularni toʻgʻrilash boʻyicha muhim resurs hisoblanadi.

TELC tarkibidagi xatolar formal va semantik turlarga boʻlinadi. Formal xatolarga soʻz yasash jarayonida notoʻgʻri qoʻshimchalar qoʻshilishi, tovush oʻzgarishlari va imlo xatolari kiradi. Semantik xatolar esa soʻzlarning notoʻgʻri tanlanishi, yaqin maʼnoli soʻzlarning aralashishi, kollokatsion muammolar va stilistik noaniqliklarni oʻz ichiga oladi. Masalan, “heavy coffee” oʻrniga “strong coffee” ishlatilishi yoki turk tilidan bevosita tarjima qilingan “do a class” kabi ifodalar notoʻgʻri deb topilgan [H. Cangir, K.Uzun, T. Can, E. Oʻguz, Ö. Faruk Kaya, 2024: 297].

TELC maxsus veb-interfeys bilan taʼminlangan boʻlib, foydalanuvchilar xatolarni koʻrish, tahlil qilish va qidirish imkoniyatiga ega. Bu korpusdan oʻqituvchilar ingliz tilini oʻrganuvchilarning eng koʻp uchraydigan xatolarini aniqlash va dars materiallarini shu asosda ishlab chiqishda foydalanishi mumkin. Talabalar esa oʻz yozma ishlarini korpus bilan taqqoslab, xatolarini mustaqil tahlil qilish imkoniyatiga ega boʻladilar. TELC ayniqsa ingliz tilini xorijiy til sifatida oʻrganayotgan turk tilida soʻzlashuvchilar uchun muhim resurs hisoblanadi va kelajakda boshqa universitetlar ishtirokida yanada kengaytirilishi mumkin.

L.Raufova rahbarligida “Oʻzbek tilining taʼlim korpusi” loyihasi asosida taʼlimiy faoliyat uchun korpus yaratilgan.

Oʻzbek tili taʼlimiy korpusini tuzilishining lingvistik asoslari boʻyicha Oʻ. Xoliyorovning tadqiqotlarida quyidagi vazifalarni oʻz ichiga oladi. Oʻzbek tili taʼlimiy korpusini tuzilishini nazariy masalalarini belgilash, oʻzbek tili taʼlimiy korpusining lingvisiyik asoslarini ochib berishdan iborat. Tadqiqotda taʼlimiy korpusning ona tili taʼlimi, oʻzbek tilini xorijiy til sifatida oʻqitishdagi afzalliklarini aniqlash, oʻzbek tili taʼlimiy korpusiga material tanlash tamoyillarini ishlab chiqish, oʻquv lugʻatlarini taʼlimiy korpusga toʻplash jarayonini aniqlash kabi ishlar amalga oshirilgan. Tadqiqotlar orqali quyidagi natijaga erishildi: oʻzbek tili taʼlimiy korpusi uchun axloqiy-tarbiyaviy matnlarni tanlash tamoyillari belgilandi va materiallar toʻplandi, oʻzbek tilidagi oʻquv lugʻatlari yagona platformaga toʻplangan va qulay qidiruv tizimiga bogʻlandi, oʻzbek tili taʼlimiy korpusining lingvistik bazasi sifatida 37200 leksamadan iborat morfoleksikon shakllantirildi, dasturiy taʼminot uchun takliflar tizimi ishlab chiqilgan hamda morfologik lugʻat tuzish tamoyillari asosida lugʻat yaratildi [Холиёров Ў, 2021: 9].

Oʻzbek tilining elektron korpusi (<http://uzbekcorpus.uz/>) Nilufar Abduraxmonova rahbarligida 2020-yil yaratilgan [Abduraxmonova N, 2023: 235].

Taʼlimiy korpus yaratish jarayonida lugʻatlarni toʻplash va tasniflashning oʻzi alohida ilmiy ahamiyatga ega. Korpusning sifatini va uning tahlil qilish imkoniyatlarini yaxshilash uchun, toʻplangan matnlarning tilshunoslik va didaktik jihatlari aniqlanishi lozim. Shu sababli, darsliklardan matnlarni toʻplash oʻquvchilarning tilni qanday oʻzlashtirayotganini, lugʻat boyligini va grammatik koʻnikmalarini aniqlashda muhim rol oʻynaydi.

Oʻquvchilar darsliklardan oʻzlashtirgan matnlar orqali tilning turli jabhalarini oʻrganadilar. Matnlarni toʻplash jarayoni, bir tomondan, oʻquvchilarning kommunikativ koʻnikmalarini rivojlantirishga, boshqa tomondan, tilni oʻrganish jarayonidagi kamchiliklarni aniqlashga yordam beradi. Bu jarayonni ilmiy nuqtai nazardan olib borish, til oʻrganishning samaradorligini oshirishga xizmat qiladi. Shu maqsadda, biz taʼlimiy korpus uchun 5-sinfdan 7-sinf gacha boʻlgan yangilangan ona tili va adabiyot darsliklaridan matnlarni toʻplab chiqdik (qarang: 1.1-rasm) va ularni korpusga qoʻshishga tayyorladik. Shu kungacha taʼlimiy korpusga 5-sinfdan 11-sinf gacha ona tili va adabiyot darsliklarining eski nashridagi matnlar kiritilgan edi.

Darsliklar oʻquvchilarga nafaqat yangi lugʻaviy birliklarni oʻzlashtirish, balki ularning nutqiy va yozma koʻnikmalarini rivojlantirishda ham yordam beradi. Ushbu amaliy ishda toʻplangan matnlar oʻquvchilarning soʻz boyligi va grammatika boʻyicha xatoliklarni tahlil qilishda, shuningdek, ularning tilni qanday oʻrganayotganini aniqlashda qoʻllaniladi. Yangilangan darsliklardan olingan matnlar korpusning yanada aniqroq va samarali tahlilini taʼminlashga imkon beradi.

*Dunyoni gullarga burkagan bola – Блокнот

Файл Правка Формат Вид Справка

Dunyoni gullarga burkagan bola

Bolakay tabiatni yaxshi ko'rarkan. Ayniqsa, bir-biridan go'zal, rang-barang xushbo'y gullarni ko'rganda beixtiyor yayrab, quvonib, bahridili ochilib ketarkan. Gullarni ona zaminning bor nafasatini o'zida jamlagan nozik bo'lagi deb bilib, ularni ko'paytirishni o'ylab yurarkan.

Kunlarning birida onasi olib kelgan gultuvaklarni ko'rib, orzusini amalga oshirish uchun birinchi qadamni tashlashga ahd qilibdi. Tuvakka gul urug'larini ekib, mehr bilan parvarish qila boshlabdi. "Gullarim gurkirab o'sib, ochilsa, uyimiz ko'rkiga ko'rk qo'shiladi,

ota-onam ham xursand bo'ladi", – degan fikr uni yanada ilhomlantiribdi.

Oradan bir necha kun o'tib, tuvakdan nafis gullar unib chiqibdi. Ularning go'zalligi barchaning havasini keltiribdi. Buni ko'rgan ota-onasi mamnun bo'lib, "Balli, o'g'lim!" – deya bolakayni maqtashibdi.

Maqtovlardan ruhlangan kichik bog'bon katta ishlarga bel bog'labdi. Qo'liga belkurakchasini olib, hovlining bir burchagiga tuvakdagi gullardan o'tqazibdi. Natijada mo'jazgina chamanzor paydo bo'libdi.

Mehnatsevar bolakay bor kuchini ishga solib, hovlisini gulzorga aylantirishga kirishibdi.

Unga ba'zan otasi, gohida onasi yordam beribdi. Jajji singlisi Rayhona ham qo'liga chelakchasini olib, gullarga suv quyibdi. Vaqt o'tib, bolakay orzu qilganidek, hovlisi gullarga burkanibdi.

Qo'ni-qo'shnilar gulzorga aylangan hovliga havas bilan qarasharkan. Mahalladosh qizaloqlar bolakaydan "Bilimlar kuni"ga gullar berishini iltimos qilishibdi. Bolakay o'zi parvarishlagan gulzordan ajoyib guldastalar yasab beribdi.

Mehnatlari ortidan rahmatlar eshitgan kichik bog'bon erta bahor uyining yonginasidan o'tgan ariq bo'yini chopa boshlabdi. Shu payt yo'l chetida yig'ilib turgan bekorchi bolakaylardan biri uni so'roqqa tutibdi:

– Ha, nima qilmoqchisan?

– Ko'chamizni gulzorga aylantiraman, – dadil javob beribdi gulsevar bolakay.

– Bo'ladigan ishni qilsang-chi?! – debdi o'yinqaroq bolalardan biri.

– Ko'cha juda katta bo'lsa, bu ishni eplolmaysan, – suhbatga qo'shilibdi koptok tepayotgan bola.

Bolalarning savol-javobini eshitib turgan qo'shni otaxon suhbatga aralashibdi:

– Men bu bolakayni yaxshi bilaman, u hovlisini gulzorga aylantirgan. Ko'chamizni ham gullarga burkashiga ishonaman.

Kichik bog'bon ter to'kib mehnat qilibdi. Dadasi uning belkurakchasini qayrab, o'tkirlab berarkan. Onasi esa bolakay va uning singlisi Rayhonaga yangi chelakchalar olib kelibdi.

Ikki-uch oyda ko'chadagi ariq yoqasi xushbo'y gullarga to'libdi. Rang-barang gullarga to'la chamanzor ko'chadan o'tayotgan har bir odamga zavq bag'ishlarkan.

1.1-rasm. 5-sinf ona tili darsligidan olingan matn

Darsliklardan iqtibosli so'zlar va lug'at birliklarini to'plash va tizimlashtirish ta'limiy korpus yaratishda muhim ilmiy jarayon hisoblanadi. Bu jarayon nafaqat o'quvchilarning yozma nutqini tahlil qilishda, balki tilni o'rganish jarayonidagi o'zgarishlarni aniqlashda ham yordam beradi. Lug'atlarni yig'ish, tasniflash va tizimlashtirish orqali o'quvchilarning til o'rganish darajasini va o'quv materiallarining samaradorligini baholash mumkin. Bunda, lug'atning kontekstdagi ishlatilishi, chastotasi va o'rganilayotgan tilga xos xususiyatlar ham hisobga olinadi.

Biz 5–7-sinf ona tili va adabiyot darsliklaridan ajratib olingan iqtibosli so'zlarni va lug'at birliklarini ta'limiy korpusga joylashtirish uchun tayyorladik (qarang: 1.2-rasm). Bu jarayon, darsliklarning o'quvchilarga tilni o'rganishda qanday yordam berishini aniqlashga imkon yaratadi. Yangi nashr etilgan 5-sinf dan 7-sinf ga qadar bo'lgan darsliklar orqali to'plangan lug'atlar, o'quvchilarning leksik boyligini oshirish va o'zlashtirish darajasini yuqori baholash imkonini beradi.

Shuningdek, lug'atlarning to'planishi va tasniflanishi, o'quvchilarning tilni o'rganish jarayonidagi xatoliklarni aniqlashda ham yordam beradi. Bu tahlil, masalan, so'zlarning noto'g'ri ishlatilishi, kam ishlatiladigan so'zlar va kerakli leksik birliklarning etishmasligini aniqlashda metodik yordamchi vosita bo'ladi. Shu sababli, lug'atlarni to'plash va tasniflash jarayoni o'qituvchilarga o'quvchilarning o'rganish jarayonini yanada samarali tashkil qilish imkoniyatini beradi.

A	B	C	D	E	F
ID	So'z	Izohi	Manbasi	Yili	Nashriyoti
1	Pediatr	bolalar shifokori	Ona tili 6-sinf	2022	Toshkent: Respublika ta'lim markazi
2	Jigarband	eng yaqin, tug'ishgan qarindosh (farzand, opa-singil, aka-uka)	Ona tili 6-sinf	2022	Toshkent: Respublika ta'lim markazi
3	Jigargo'sha	farzand	Ona tili 6-sinf	2022	Toshkent: Respublika ta'lim markazi
4	Jigarpora	jigarband, jigarinning bir parchasi, yaqinim	Ona tili 6-sinf	2022	Toshkent: Respublika ta'lim markazi
5	Jigarrang	qizg'ish, qizil, jigar rangiga o'xshash	Ona tili 6-sinf	2022	Toshkent: Respublika ta'lim markazi
6	Jigarso'xta	jigari kuygan, yongan. Og'ir musibat chekib ezilgan, ko'ngli yarim bo'lib qolgan	Ona tili 6-sinf	2022	Toshkent: Respublika ta'lim markazi
7	Jigari kuygan, yongan	Og'ir musibat chekib ezilgan, ko'ngli yarim bo'lib qolgan	Ona tili 6-sinf	2022	Toshkent: Respublika ta'lim markazi
8		holidan xabar olmoq, tashrif buyurmoq, o'tirish			

2.4-rasm. 6-sinf ona tili darsligidan to'plangan lug'atlar

Yuqoridagi rasmda 6-sinf ona tili darslikdagi mavjud lug'atlarni va iqtibosli so'zlarni to'plab, ularni elektron jadval shaklida tizimlashtirdik. Har bir so'z uchun individual identifikatsiya raqami (ID), so'zning o'zi, uning matndagi izohi, manba sifatida darslik nomi, sinfi, nashriyoti va chop etilgan yili kiritildi. Bu tizimlashtirish metodik jihatdan muhim bo'lib, har bir so'zning tahlilini osonlashtiradi va o'quvchilarning lug'at boyligi haqida aniq tasavvur hosil qiladi.

Shu asosida 5-sinfdan 7-sinfgacha bo'lgan yangi nashrdagi darsliklarda berilgan lug'atlar va iqtibosli so'zlarni korpusga joylashtirish uchun tayyorladik.

Xulosa

Ta'limiy korpuslar lingvistik tadqiqotlar va til o'rgatish jarayonida muhim vositadir. Ular o'quvchilarning til o'zlashtirish darajasini aniqlash, grammatik xatolarni tahlil qilish va samarali o'quv materiallarini ishlab chiqishda katta rol o'ynaydi. Ingliz, turk va o'zbek tillaridagi ta'limiy korpuslar til o'rganuvchilarining yozma ishlari asosida yaratilgan va xatolarni aniqlashda samarali vosita bo'lib xizmat qiladi.

O'zbek tili ta'limiy korpusi yaratilishi jarayoni ilmiy va amaliy jihatlarni o'z ichiga oladi. Matn tanlash va lug'atlarni tizimlashtirishning ilmiy asoslari hamda amaliy ahamiyati ta'limiy korpuslarni samarali foydalanishga imkon beradi. Bu korpuslar, til o'rgatish jarayonida grammatik xatolarni tahlil qilish va tilni o'rganishning samarali metodologiyalarini ishlab chiqishda yordam beradi.

Amaliy jihatdan, ta'limiy korpuslar til o'rganish jarayonini tizimli va ilmiy asosda olib borishga imkon yaratadi, bu esa o'quvchilarning til o'zlashtirish darajasini aniqlash va samarali o'quv materiallarini ishlab chiqishda muhim ahamiyatga ega.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Korpus linvistikasi, Toshkent, 2023. – B. 56
2. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
3. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).

4. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
5. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus. uz. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.
6. Amirbekova A.B. va boshqalar. Milliy korpusda matnlarni avtomatlashtirish usullari. 2023
7. Cem Can. A learner corpus-based study on verb errors of Turkish EFL learners. //Journal of Education and Training Studies. 2017. – B. 13
8. Erdem Akbas. Corpus of Turkish undergraduate writers of English. 2022.
9. H. Cangır, K.Uzun, T. Can, E. Oğuz, Ö. Faruk Kaya, Turkish-English Learner Corpus and its Web-interface. // Dilbilim Araştırmaları Dergisi. 2024. – P.279-307.
10. Mauranen Anna and Elina Ranta. English as a Lingua Franca: Studies and Findings. Newcastle: Cambridge and Scholars, 2009. 562.
11. Sidiropoulos N, Hatzigeorgiu N. Construction of a web based corpus for educational uses. //Institute for language and speech processing, 2011. – B. 4-5
12. Холиёров Ў. М. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари. Филол. фан.б.фалс. док.дисс, Термиз, 2021. – С. 9-15

COMBINING CORPUS LINGUISTICS WITH CRITICAL DISCOURSE ANALYSIS

Asrorova Nargiza

Teacher at Tashkent International

University of Financial Management Technologies

E-mail: nibotova@gmail.com

Abstract. This study explores the integration of Corpus Linguistics (CL) with Critical Discourse Analysis (CDA) to enhance the reliability and interpretive depth of discourse studies. By combining the quantitative strengths of CL with the qualitative focus of CDA, the study suggests methods to uncover underlying discourse features and discursive patterns in large-scale textual data. The paper outlines key methodological frameworks that facilitate this interdisciplinary approach, including keyword analysis, concordance patterns, and collocation networks, alongside critical interpretation rooted in socio-political contexts. Through illustrative examples, it demonstrates how corpus tools can identify linguistic patterns that might otherwise go unnoticed, supporting a more systematic and transparent analysis of discourse. The synergy between CL and CDA not only enhances analytical precision but also broadens the scope of discourse analysis, making it more replicable and data-driven. This combined approach is particularly useful for investigating second language, media discourse, and social representations in contemporary communication.

Keywords: Applied Linguistics, Corpus Linguistics, Critical Discourse Analysis, Sociolinguistics, Concordance Patterns

Annotatsiya. Ushbu tadqiqot tanqidiy diskurs tahlili (CDA) va korpus lingvistikasini (CL) uyg'unlashtirish yondashuvi orqali diskurs tahlilining ilmiy asoslanganligini va interpretativ chuqurligini oshirishni maqsad qilgan. Tadqiqot tanqidiy diskurs tahlilining miqdoriy tahlil salohiyatini korpus lingvistikasining tavsifiy yondashuvi bilan birga qo'llaganda, keng ko'lamli matnlarda diskurs xususiyatlari va jihatlarini o'rganadi. Maqolada ushbu fanlararo metodologiyaga asos bo'ladigan usullar – kalit so'zlar tahlili, konkordans (muvofiqlik) va so'z birikmalari tahlil qilinadi. Mazkur maqolada keltirilgan misollar korpus vositalari orqali ko'pincha e'tibordan chetda qoladigan til birliklarini aniqlash imkonini beradi. Tanqidiy diskurs tahlili va korpus lingvistikasi integratsiyasi nafaqat tahlil aniqligini oshiradi, balki diskurs tahlilini tizimli holga keltiradi, ilgari mumkin bo'lmagan jihatlarini ochish imkonini beradi. Ushbu integratsiyalashgan yondashuv ayniqsa ikkinchi til, media diskursi va ijtimoiy tasvirlarni o'rganishda dolzarb ahamiyatga ega.

Kalit so'zlar: Amaliy tilshunoslik, korpus lingvistikasi, tanqidiy nutq tahlili, sotsiolingvistika, konkordanslar.

Аннотация. В данном исследовании рассматривается интеграция корпусной лингвистики (CL) и критического дискурс-анализа (CDA) с целью повышения обоснованности и интерпретативной глубины дискурс-исследований. Объединяя количественные возможности корпусной лингвистики с качественным фокусом критического дискурс-анализа, исследование выявляет скрытые дискурсивные особенности и паттерны в масштабных текстовых данных. В статье представлены ключевые методологические подходы, способствующие реализации этого междисциплинарного подхода, включая анализ ключевых слов, паттерны конкорданса и сети коллокаций, а также критическую интерпретацию, основанную на социально-политическом контексте. На основе иллюстративных примеров демонстрируется, как корпусные инструменты позволяют выявлять языковые закономерности, которые в противном случае могли бы остаться незамеченными, обеспечивая более системный и прозрачный анализ дискурса. Синергия между CL и CDA не только усиливает аналитическую точность, но и расширяет возможности дискурсивного анализа, делая его более воспроизводимым и ориентированным на данные. Такой комбинированный подход

особенно полезен для исследования иностранного языка, медиадискурса и социальных репрезентаций в современной коммуникации.

Ключевые слова: Прикладная лингвистика, корпусная лингвистика, критический дискурс-анализ, социолингвистика, паттерны конкорданса.

Introduction

Variety of theoretical frameworks and methodological approaches have been applied in linguistic research to link more than one aspect of language studies. One of such combinations is the interdisciplinary approach to combine corpus linguistics (CL) with critical discourse analysis (CDA) which offers innovation to the field⁶⁸.

Corpus Linguistics (CL) is a data-driven approach to language analysis that relies on corpora as its main source of evidence. It involves identifying linguistic patterns, trends, and associations within large text datasets, enabling generalizations about language use⁶⁹. Core methods include examining word lists, keywords, collocations, and concordances, often supported by statistical analysis. Although largely quantitative, CL focuses on local textual contexts, especially through concordance analysis—studying repeated word occurrences in varied environments⁷⁰. Beyond surface-level analysis, CL is increasingly used in discourse studies to explore how specific linguistic features function across texts and language varieties.

Critical Discourse Analysis (CDA) treats language as both a discourse and a form of social practice, emphasizing the connection between language and ideology⁷¹. Influenced by Halliday's view of language as deeply embedded in its sociolinguistic context⁷², CDA explores how texts reflect and shape power relations within society. Its analytical process typically involves three interconnected stages: describing linguistic features in the text, interpreting how the text interacts with its context, and explaining the broader social implications of this interaction. Through this framework, CDA uncovers how language both reflects and constructs social realities⁷³.

CDA is mainly criticized for its qualitative nature lacking methodology for sampling, collecting texts and comparing textual features. As a solution, combination of corpus linguistics and critical discourse analysis was suggested. Several studies have successfully combined Critical Discourse Analysis (CDA) and Corpus Linguistics (CL)⁷⁴ by examining word frequencies, collocations, and syntactic patterns. These include analyses of pronoun use in sexist language, transitivity in educational texts, and media representations of women, the EU, and ethnicity. Other studies explored environmental discourse, political rhetoric, and press representations across cultures. One example focuses on modal certainty in political interviews, analyzing various linguistic forms that convey speaker commitment. Another study applied corpus-based CDA to a single academic paper, examining frequent terms and their concordances to reveal deeper interpretive insights into discourse construction and ideological framing.

Methodology

Two primary types of corpus-based discourse analysis are corpus-assisted critical discourse analysis and corpus-informed discourse analysis. Baker et al.⁷⁵ use a corpus-assisted CDA

⁶⁸ Baker, P., Gabrielatos, C., Khosravini, M., Krzyżanowski, M., McNery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse and Society*, 19(3), 273–306

⁶⁹ Teubert, W., & Krishnamurthy, R. (2007). General introduction. In W. Teubert & R. Krishnamurthy (Eds.), *Corpus linguistics: Critical concepts in linguistics* (pp. 1–37). London, England: Routledge

⁷⁰ Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford, England: Oxford University Press

⁷¹ Fairclough, N. (2001). *Language and power* (2nd ed.). London, England: Longman.

⁷² Halliday, M. A. K. (1985). *An introduction to functional grammar*. London, England: Edward Arnold.

⁷³ The same study

⁷⁴ Flowerdew, J. (2004). Identity politics and Hong Kong's return to Chinese sovereignty: Analyzing the discourse of Hong Kong's first chief executive. *Journal of Pragmatics*, 41(11), 2333–44

⁷⁵ Baker, P., Gabrielatos, C., Khosravini, M., Krzyżanowski, M., McNery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse and Society*, 19(3), 273–306

approach to study how refugees, asylum seekers, immigrants, and migrants (RASIM) are portrayed in the British press from 1996 to 2005. They compile a 140-million-word corpus and explore keyword frequency and collocational patterns around the core terms. After identifying statistically significant trends, they conduct qualitative analyses of selected texts, drawing on broader socio-political contexts related to migration. This method allows them to link linguistic patterns with ideological representations in media discourse, demonstrating how CL can enrich CDA through both breadth and depth of analysis.

Galasinski and Marley⁷⁶ adopt a corpus-informed approach by comparing British and Polish press corpora to examine representations of agency in foreign news. They begin with a quantitative content analysis to detect media trends, which then guide a qualitative CDA for deeper interpretation. Their findings reveal that both countries' media tend to frame foreign coverage through the lenses of business and politics—business being more prominent in British news and politics in Polish news. Additionally, British media show a complex, often contradictory portrayal of Russia, which the qualitative analysis explores further in terms of negative discourse patterns⁷⁷.

Conclusion

An increasing number of studies have successfully merged Corpus Linguistics (CL) with Critical Discourse Analysis (CDA) to examine pressing social issues, processes, and phenomena by leveraging the unique strengths of each discipline. While most CDA research has focused on specialized corpora—particularly those related to media and political discourse—future research could expand into new areas, such as educational feedback⁷⁸, academic classroom interactions⁷⁹ or business discourse (Koller, 2005). Additionally, CL can be aligned with diverse theoretical CDA frameworks, including social semiotics⁸⁰ and the discourse-historical method⁸¹, enhancing both depth and methodological diversity.

To combine Corpus Linguistics (CL) and Critical Discourse Analysis (CDA), start by constructing a relevant corpus, ensuring it reflects the discourse under study. Perform quantitative CL analyses, including word frequency, concordance, and collocational patterns. Then, apply qualitative CDA methods to interpret these findings in social, political, and ideological contexts, focusing on power dynamics and discourse strategies. Integrate insights from other fields (e.g., sociocognitive studies, intercultural rhetoric) for deeper analysis. Use data triangulation to enhance reliability and validate results. Present findings with a mix of quantitative results and critical interpretations, using visual tools for clarity.

References:

1. Baker, P., Gabrielatos, C., Khosravini, M., Krzyżanowski, M., McNery, T., &
2. Csomay, E. (2006). Academic talk in American university classrooms: Crossing the boundaries of oral-literate discourse? *Journal of English for Academic Purposes*, 5, 117–35
3. Fairclough, N. (2001). *Language and power* (2nd ed.). London, England: Longman.
4. Flowerdew, J. (2004). Identity politics and Hong Kong's return to Chinese sovereignty: Analyzing the discourse of Hong Kong's first chief executive. *Journal of Pragmatics*, 41(11), 2333–44
5. Galasinski, D., & Marley, C. (1998). Agency in foreign news: A linguistic complement of a content analytical study. *Journal of Pragmatics*, 30, 565–87

⁷⁶ Galasinski, D., & Marley, C. (1998). Agency in foreign news: A linguistic complement of a content analytical study. *Journal of Pragmatics*, 30, 565–87

⁷⁷ The same source

⁷⁸ Hyatt, D. F. (2005). "Yes, a very good point!" A critical genre analysis of a corpus of feedback commentaries on Master of Education assignments. *Teaching in Higher Education*, 10(3), 339–53

⁷⁹ Csomay, E. (2006). Academic talk in American university classrooms: Crossing the boundaries of oral-literate discourse? *Journal of English for Academic Purposes*, 5, 117–35

⁸⁰ The same source

⁸¹ Wodak, R. (2001). The discourse historical approach. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (pp. 63–94). London: Sage.

6. Halliday, M. A. K. (1985). An introduction to functional grammar. London, England: Edward Arnold.
7. Hyatt, D. F. (2005). “Yes, a very good point!” A critical genre analysis of a corpus of feedback commentaries on Master of Education assignments. *Teaching in Higher Education*, 10(3), 339–53
8. Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford, England: Oxford University Press
9. Teubert, W., & Krishnamurthy, R. (2007). General introduction. In W. Teubert & R. Krishnamurthy (Eds.), *Corpus linguistics: Critical concepts in linguistics* (pp. 1–37). London, England: Routledge
10. Wodak, R. (2001). The discourse historical approach. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (pp. 63–94). London: Sage.

AMALIY TILSHUNOSLIKDA LINGVISTIK EKSPERTIZA VA UNING AHAMIYATI

O'rozov Jumanazar Norqizilovich

Mirzo Ulug'bek nomidagi

O'zbekiston Milliy universiteti katta o'qituvchisi,
filologiya fanlari bo'yicha falsafa doktori (PhD)

Annotatsiya: Hozirgi kunga kelib fan va texnikaning jadal rivojlanishi fanlararo integratsiyaning kuchayishi fan tarmoqlarida o'zaro ta'sir, muammo va masalalar yechimida hamkorlikning yuzaga kelishiga omil bo'ldi. Jamiyatda ijtimoiy fanlarning o'zaro bog'liqligi, fanlararo aloqalarning rivojlanishi, xususan, til va huquq fanlarining bir-birini taqozo qilishi, yuridik materiallarni til nuqtayi nazaridan tadqiq qilishga ehtiyoj yangi yo'nalish – lingvistik ekspertizaning shakllanishiga olib keldi. Maqolada lingvistik ekspertiza amaliy tilshunoslikning bir yo'nalishi sifatida o'rganilishi, ular haqida olimlarning nazariyalari haqida ma'lumot berilgan. Lingvistik ekspertiza boshqa sud ekspertizalari bilan chambarchas bog'likligi haqida ma'lumot berilgan.

Kalit so'zlar: lingvistik ekspertiza, amaliy tilshunoslik, sud lingvistikasi, sud lingvistik ekspertizasi

Abstract: To date, the rapid development of science and technology, the strengthening of interdisciplinary integration, has become a factor in the emergence of interaction in scientific fields, cooperation in solving problems and issues. The interdependence of social sciences in society, the development of interdisciplinary relations, in particular, the mutual dependence of linguistics and law, the need to study legal materials from a linguistic point of view have led to the formation of a new direction - linguistic expertise. The article provides information on the study of linguistic expertise as a direction of applied linguistics, the theories of scientists about them. Information is provided on the close relationship of linguistic expertise with other forensic expertise.

Keywords: linguistic expertise, applied linguistics, forensic linguistics, forensic linguistic expertise

Аннотация: На сегодняшний день бурное развитие науки и техники, активизация междисциплинарной интеграции привели к возникновению взаимодействия в научных сетях, сотрудничества в решении проблем и вопросов. Взаимосвязь социальных наук в обществе, развитие междисциплинарных связей, в частности взаимообусловленность языковых и юридических наук, необходимость исследования правового материала с точки зрения языка привели к формированию нового направления лингвистической экспертизы. В статье представлены сведения об изучении лингвистических экспертиз как отрасли прикладной лингвистики и теории ученых о них. Лингвистическая экспертиза тесно связана с другими судебно-медицинскими экспертизами.

Ключевые слова: лингвистическая экспертиза, прикладная лингвистика, судебная лингвистика, судебная лингвистическая экспертиза.

Bugungi kunda mamlakatimizning yanada demokratlashuvi, so'z erkinligi, ommaviy-axborot vositalarining juda faollashuvi jarayonlarida matnni lingvistik aspektida ekspertizadan o'tkazish ishlariga talab kuchayib bormoqda. Buni lingvistik ekspertizani amalga oshirish uchun tergov, surishtiruv organlari tomonidan tilshunoslik ilmiy-tadqiqot muassasalariga, tilshunos olimlarga bo'layotgan murojaatlarning ko'pligidan ham ko'rish mumkin. Jumaladan, mualliflik huquqi, plagiat va qalbaki bosma materiallarga qarshi kurashish, haqorat insonning sha'ni, qadr-qimmatini va kasbiy obro'sini himoya qilish sohasida ko'plab muammolar paydo bo'lmoqda. Shu nuqtai nazardan, bu sohani respublikamiz miqyosida rivojlantirish, uning amaliy ahamiyatini

kengaytirish, xususan, lingvistik ekspertiza faoliyatini takamillashtirish dolzarb ahamiyat kasb etadi.

Lingvistik ekspertiza ahamiyatining keskin oshishi, dunyoda unga bo'lgan ehtiyojning kengayishi, xususan, qayta qurish, so'z erkinligi, reklama, bozor munosabatlarining rivojlanishi, milliy muammolar - ommaviy axborot vositalari matnlari bilan bog'liq ko'plab sud ishlari bilan bog'liq, ommaviy nutqlar, reklama, tijorat nomlari, identifikatsiya tahlillari va diagnostika (matn muallifining belgilari) zarur.

“Lingvistik ekspertiza” tushunchasi lotincha “linguistic expertise” dan olingan bo'lib, o'zbek tilida *lingvistika – tilshunoslik, ekspertiza – tekshirish, tahlil qilish, sinov, tajriba* degan ma'nolarni beradi.

Muayyan bir nutq turini (og'zaki yoki yozma nutqni), matnni shu soha mutaxassisi tomonidan tilshunoslik qonun-qoidalariga asosan shakliy va mazmunijihatdan o'rganilishi, tekshirilishi, tahlil qilinishi, tadqiq qilinishi, matnning davlat tili, adabiy til, vazifaviy uslublar qoida va talablariga muvofiqligining aniqlanishi lingvistik ekspertiza deyiladi.

Xorijiy ilmiy adabiyotlarda lingvistik ekspertiza dastlab “sud lingvistikasi” (forensic linguistics) sifatida yuritilgan. Bu nomning kelib chiqishi Britaniyalik yozuvchi va olim Yan Svartvik bilan bog'liq.

Yan Svartvik 1968-yilda London shahrining Kensington qismida o'z xotini va bolalarini o'ldirganlik jinoyatida ayblangan Timot Evansning sudi ishi materiallarini lingvistik ekspertizadan o'tkazadi va uning aybsizligini isbotlab beradi. Ya'ni, Timot Evansning sud jarayonida bergan barcha yozma ko'rsatmalarini qiyosiy taqqoslab, aybini bo'yniga olganligi haqidagi hujjatning til va uslubiga qarab, mazkur hujjatni Timot Evans yozmaganligini, politsiya xodimlarining o'zlari yozganligini isbotlab beradi. [Плотникова, С. Н., 2006:2]

Vaholanki, ungacha Timot Evansga nisbatan sud hukmi e'lon qilinib, u qotil sifatida qatl etilgan edi va o'limidan so'ng Yan Svartvik lingvistik ekspertizasi natijalariga ko'ra oqlanadi.

Bu voqea-hodisa keng jamoatchilik o'rtasida katta shov-shuvlarga sabab bo'ladi. O'z navbatda, yurilingvistika fanining sud lingvistikasi sifatida rivojlanishiga sabab bo'ladi. Olimlar tomonidan mazkur fanga bo'lgan e'tiborni yanada oshiradi.

Lingvistik ekspertizaning o'rganish obyekti aniqlashda ham turli qarashlar mavjud. Jumladan, N.D.Golev til va huquq o'rtasidagi munosabatlardan kelib chiqqan holda lingvistik ekspertizaning o'rganish obyekti sifatida quyidagilarni ajratib ko'rsatadi:

– “til yuridik nuqtayi nazardan turli ziddiyatli holatlarda tartibga solish vositasi sifatida qo'llanadi;

– shuningdek, til tartibga solishni amalga oshiradigan vosita sifatida ishlaydi”. [Голев Н.Д., 1999:1]

Bizning fikrimizcha, N.D.Golevning bu nazariyasi orqali tilning yurisprudensiyadagi vazifasi oydinlashsa, ikkinchi ta'rif tilning nafaqat yuridik jihatdan, balki istalgan bahsli matnlarni tartibga solishni amalga oshiradigan jarayon ekanligini anglatadi.

Aytish joizki, yaqin vaqtgacha sud-tergov jarayoni bilan bog'liq matnlarni lingvistik ekspertizadan o'tkazish ekspert muassasalaridan tashqarida tilshunos olimlar tomonidan amalga oshiriladi. Amaliyotda esa filolog olimlar qanday tamoyillar asosida ekspertizani amalga oshirish, bu jarayonda asosiy e'tiborni nimaga qaratish kabi masalalarda qiyinchiliklar duch kelinar edi. Professional mutaxassis bo'lmagan filologlarning xulosalari ko'pincha tasdiqlanmagan versiyalar, taxminlar, gipotezalar ilmiy muhokama uchun maqbul bo'lishi mumkin bo'lgan, ammo ekspert xulosasi sifatida mos kelmaydigan subyektiv fikrlardan iborat bo'ladi hamda, o'z navbatida, ishonchliligi ilmiy tasdiqlashni talab qiladigan tekshirilmagan lingvistik usullardan foydalanilgan. Natijada bunday xulosalar sud qarorlarida ijobiy baholanmas yoki sudlar tomonidan umuman dalil sifatida qabul qilinmagan.

Amaldagi protsessual qonun hujjatlari talablariga muvofiq lingvistik ekspertizalarni ishlab chiqarish uchun ijtimoiy buyurtma yuqorida ko'rsatilgan toifalardagi ishlarni sudda ko'rib chiqish uchun professional lingvistik ekspertni obyektiv ravishda talab qiladi.

Aytish kerakki, “lingvistik ekspertizasi” termini dastlab jamoaviy termin sifatida qaralar edi. Xususan, “tekstologik”, “filologik”, “stilistik”, “semasiologik” ekspertizalar va boshqalar nomlari ishlatilgan, ammo nutq faoliyati bilan bog‘liq turli xil ekspert muammolarini amalga oshiradi.

Lingvistik ekspertizaning obyekti – ixtiyoriy moddiy axborot tashuvchida qayd etilgan til va nutq birligi, og‘zaki, yozma, elektron matnlar hisoblanadi.

Lingvistik ekspertizaning predmeti – tilshunoslik sohasida maxsus bilim talab qiladigan masalalarni hal etish orqali aniq ish bo‘yicha isbotlanishi lozim bo‘lgan fakt va holatlarni aniqlashdan iborat.

Lingvistik ekspertiza maqsadi – maxsus filologik bilimlarga asoslanib, aniq ish bo‘yicha isbotlanishi kerak bo‘lgan holatlarni aniqlashdir.

Bugun biz lingvistik ekspertiza doirasida hal qilinadigan quyidagi tipik vazifalar haqida gapirishimiz mumkin:

- 1) munozarali matnni, bayonotni, uning semantik mazmunini aniqlash maqsadida tadqiq qilish;
- 2) munozarali matn janr, kompozitsion yoki leksiko-grammatik ifoda shakli nuqtayi nazaridan o‘rganish;
- 3) tijorat nomlarini (savdo nomlari, tovar belgilari, domen nomlari) tadqiq qilish, ularning o‘ziga xosligi, individualligi, yangiligi, chalkashlik darajasini aniqlash;
- 4) normativ-huquqiy hujjatlar loyihalarini davlat tili qoidalariga moslash, terminologik, uslubiy nuqtayi nazardan tekshirish.

Ushbu vazifalar ro‘yxati to‘liq emas va sud amaliyotining so‘rovlariga muvofiq kengaytirilishi mumkin.

Xulosa qilib shuni aytish mumkinki, O‘zbekistonda lingvistik ekspertiza sohasini ilmiy, amaliy va ta’lim sohasida rivojlantirish dolzarb vazifa hisoblanadi. Bu esa davlat va jamiyat taraqqiyotida muhim ahamiyat kasb yetadi. Hozirgi axborot urushlari sharoitida ommaviy axborot vositalarida so‘z erkinligini suiiste‘mol qilishga qarshi kurash zarur bo‘lib qolmoqda. Aytish joizki, ba’zida ekspert tilshunosning xulosasisiz aybni qonuniy ravishda isbotlash imkonsiz bo‘lib qoladi. Shuning uchun bugungi kunda mazkur sohani rivojlantirish dolzarblik kasb etmoqda.

Fikrimizcha, bu sohani dunyo ilm-fani darajasida rivojlantirish hamda ilmiy-amaliy ahamiyatini oshirish uchun quyidagi masalalarga e’tibor qaratish zarur:

Birinchidan, Respublikadagi barcha sud ekspertizalari qatori lingvistik ekspertizani ham mustaqil sohaga aylantirish lozim. Chunki uning konsepsiyasi matn bilan bog‘liq barcha ekspertizalarni o‘z ichiga oladi. Shu bilan birga u til bilan bog‘liq bo‘lgan barcha ekspertizalarni birlashtirishi kerak. Xususan, mualliflik, qo‘lyozma, hujjatlashtirish, sha’n va qadr-qimmat tushunchasi bilan bog‘liq ekspertiza, haqorat va hokazo. Unga qo‘shimcha qilib, grafologiya bo‘yicha tasdiqlangan bilimlarni ham jalb qilish mumkin.

Ikkinchidan, barcha huquqni muhofaza qiluvchi organlar turli ko‘rinishdagi kriminalistik matnlar, ijtimoiy-siyosiy jihatdan xavfli, shubhali holatlarni (turli siyosiy nutqlar, reklamalar, nomlar) aniqlash uchun lingvistik hamda huquqiy bilimlarga ega bo‘lishi uchun ularning shu soha bo‘yicha malaka, bilim va ko‘nikmalarini oshirish maqsadga muvofiq.

Uchinchidan, lingvistik ekspertiza tergov jarayonida xolis xulosa beruvchi sohadir. Shu boisdan mamlakatimizda lingvist ekspert tayyorlash, malakasini oshirish mexanizmlarini ishlab chiqish, lingvistik ekspertiza markazlar faoliyatini yo‘lga qo‘yish lozim.

To‘rtinchidan, hozirgi vaqtda turli sohalar mutaxassislari ishtirok etadigan kompleks tekshiruvlar mavjud. Ularda lingvistik ekspertizaga alohida urg‘u berilishi kerak.

Foydalanilgan adabiyotlar ro‘yxati:

1. Голев Н.Д. Юридический аспект языковингвистическом освещении. Юрислингвистика. Проблемы и перспективы: Межвуз.сб. научных трудов / Под. ред. Н.Д.Голев. – Барнаул: Изд-во Алт. ун-та, 1999. – С. 7–38.

2. Плотникова, С. Н. Лекции по судебной фонетике / С. Н. Плотникова. - Иркутск : ИГЛУ, 2006. - 58 с
3. Кусов, Г.В. Судебная лингвистическая экспертиза: понятие, принципы, процессуальное значение: монография. – Краснодар: Дом - ЮГ, 2012. – 284с.
4. Кўчимов Ш. Ўзбек тилшунослигида фанлараро интеграция ва унинг давлат ҳамда жамият тараққиётига таъсири (тил ва ҳуқуқ фанлари мисолида) // Ўзбек тили ва адабиёти журнали, 2022 йил, 1 сон, – Б. 59-62.
5. O'rozov J.N. O'zbek tili lingvokriminalistikasi. T., Bookmany print, 2023. 137 b.

BADIIY MATN LINGVOPOETIKASIDA G'AYRIODATIIY BIRIKMALAR

Jamolitdinova Dilnoza Mirhojiddinova
QDPI professori, f.f.d., DSc

Akbarova Munira Shavkatovna
QDPI o'qituvchisi, f.f.b.f.d., PhD
E-mail: akbarovam1989@mail.ru

Annotatsiya: mazkur maqolada badiiy matnda tilning bajaradigan kommunikativ vazifasi, estetik vazifasi, amaliy vazifasi haqida so'z yuritiladi, badiiy matnni lingvopoetik tahlili uchun badiiy qimmat kasb etuvchi g'ayriodatiiy birikmalardan namunalar keltirilgan hamda badiiy lingvopoetik vazifasi tahlil qilingan.

Kalit so'zlar: tilning kommunikativ vazifasi, estetik vazifasi, lingvopoetika, amaliy vazifa, badiiy matn lingvopoetikasi.

Abstract: this article discusses the communicative function, aesthetic function, and practical function of language in a literary text, provides examples of unusual combinations that give artistic value for the linguopoetic analysis of a literary text, and analyzes the linguopoetic function of art.

Keywords: communicative function of language, aesthetic function, linguopoetics, practical function, linguopoetics of a literary text.

Аннотация: В статье рассматриваются коммуникативная, эстетическая и практическая функции языка в художественном тексте. Приводятся примеры необычных сочетаний, придающих художественную ценность лингвопоэтическому анализу литературного текста, анализируется лингвопоэтическая функция литературы.

Ключевые слова: коммуникативная функция языка, эстетическая функция, лингвопоэтика, практическая функция, лингвопоэтика художественного текста.

Til kishilik jamiyatining mohiyatini, inson ruhiyati va uning faoliyatini turlicha qarashlarda tadqiq etuvchi barcha fanlarning eng muhim markaziy, ayni damda, eng murakkab izlanish manbayi bo'lgan o'ziga xos hodisa hisoblanadi. Shu murakkab hodisaning yuzaga chiqishi, reallashuvi, namoyon bo'lishi jarayonida yaxlit bir butunlik sifatida nutq oqimi yaratiladi. Uning natijasida badiiy nutq yuzaga keladi. Mana shu badiiy nutqni bo'laklarga ajratish, har bir bo'lak yoki bo'lakchani matndan ayro holda tasavvur qila olish, shuningdek, bo'laklar o'rtasida hosil bo'ladigan munosabatlarni idrok etishdek bir jarayon kabi o'rgana boshlaguniga qadar inson ongi ancha uzoq hamda murakkab va mashaqqatli rivojlanish yo'lidan yurib o'tdi.

Badiiy bo'lmagan nutqda yaratilgan matnda asosiy maqsad o'quvchiga kommunikativ ta'sir ko'rsatish bo'lsa, badiiy matnda esa muallifning asosiy maqsadi kitobxonga estetik ta'sir ko'rsatishdan iborat bo'ladi. Badiiy va badiiy bo'lmagan matnlar o'zaro tilning estetik vazifasi namoyon bo'lishiga ko'ra farqlanadi. "Zero, vazifasiz til bo'lmaydi, materiya harakatsiz mavjud bo'la olmaganidek, til ham vazifadan tashqarida yashay olmaydi, demakki, vazifa tilning mavjudlik omilidir"[Авдорин В.А., 1975. 34].

Shu o'rinda professor N.Mahmudovning mana bu fikrlari diqqatga molik: "Til fikr ifodalash, dunyoni bilish, bilim-tajribalarni to'plash, saqlash va keyingi avlodlarga yetkazish, milliy-ruhiy munosabatlarni aks ettirish, badiiylik kategoriyalarini voqelantirish kabi bir qancha vazifalarni bajaradi. Tilni faqat va faqat kishilar o'rtasidagi aloqa vositasi sifatida talqin etish insonning tabiiy tilini, bu murakkab va muhtasham hodisani, eng kami, jo'nlashtirishdan, aniq bir milliy qiyofa yoki milliy-ruhiy zamindan mosuvo bo'lgan sun'iy tilga (masalan, Esperanto tiliga) tenglashtirishdan, yo'l harakatini tartibga solish maqsadida yaratilgan shartli "til"ga

baravarlash tirishdan boshqa narsa emas... Holbuki, odamlar til vositasida tuyg‘u va kechinmalari, quvonch va qayg‘ulari, hayrat va hayronliklari, qalbdagi huzurlari kabi xilma-xil sezgilarini ham ifodalaydilarki, bular hamisha ham sof kommunikativ maqsadlarni ko‘zda tutmaydi... Estetik vazifa tilning asosiy kommunikativ vazifasini qulaylashtiradi, uning imkoniyatlarini boyitadi. Til va nutq ayni shu estetik vazifasi bilan ifodalilik, ta’sir quvvatini namoyon etadi”[Махмудов Н.М., 2006. 47].

Har qanday badiiy matnda til kommunikativ vazifani ham bajaradi, biroq estetik vazifa, albatta, birinchi o‘rinda turadi, yetakchi hisoblanadi. Shu bois lingvistik manbalarda bu holat quyidagicha izohlanadi: “Badiiy matn har qanday nobadiiy matndan farqli o‘laroq alohida vazifani – kommunikativ vazifa bilan murakkab o‘zaro aloqadorlikda namoyon bo‘luvchi va matnning o‘ziga xos qurilishida hal qiluvchi omil hisoblanuvchi estetik vazifani bajaradi”[Гореликова М.И., 1989. 5]. A.I.Gorshkov o‘rinli ta’kidlaganidek, “estetik vazifani tilning boshqa “amaliy” vazifalaridan farqlaydigan (ammo ajratib qo‘ymaydigan) narsa – bu nutqiy ifoda shaklining u yoki bu mazmunni berishgagina emas, balki shu ifodaning o‘ziga, tilning o‘zidagi go‘zallikni sezish, undan zavq olishga imkon beruvchi mukammallikka yo‘naltirilganligidir. Agar tilning “amaliy” vazifalari axborotni aniq, ravshan va tushunarli ifodalash maqsadi bilan so‘z ustida ishlashni taqozo etsa, tilning estetik vazifasi bundan tashqari o‘quvchi va tinglovchiga ifodaning o‘zidagi go‘zallikni ochib berish maqsadi bilan so‘z ustida ishlashni talab etadi”[Горшков А.И., 2001. 324].

Badiiy matnda fikrning ifodalanishi, ochiq va yashirin ifodalanish usullarini o‘rgangan turk tilshunosi J.Gariberning mazkur fikrlari e’tiborga molik: “...Muammo fikrning estetik jihat bilan qay tarzda bir butunlik hosil qilishi, tilga qanday ko‘chishi masalasidir. San’at asari ilmiy asardan, adabiyot tadqiqotchisining yondashuvi boshqa ilmiy yondashuvlardan ana shu nuqtada farq qiladi. Adabiyot tadqiqotchisi badiiy asardagi fikrni ham estetik uzvning bir qismi sifatida baholashga majburdir”[Gariper C., 2008. 159].

Badiiy matn so‘z ijodkorining voqelikni, olamni obrazli-estetik tarzda idrok qilishi natijasida yuzaga keladi. Badiiy matnning yuzaga kelishi ham, matnning o‘zi ham, asarni o‘quvchi tomonidan uni badiiy butunlik sifatida anglay olinishi, tasavvur doirasida idrok etilishi ham o‘ziga xos murakkab estetik faoliyatdir. Z.I.Xovanskaya bu turdagi estetik faoliyatning tarkibiy qismlarini quyidagicha sifatlaydi: “1) muallif niyatining va ijodkor ongida voqelikning estetik in’ikosi natijasida badiiy obrazlarning paydo bo‘lishi; 2) badiiy obrazlarning estetik materialda gavdalanishi, ya’ni ularning tegishli san’at turi vositalari orqali moddiylashuvi va estetik faoliyat mahsuli – san’at asarining yuzaga kelishi; 3) san’at asarining omma tomonidan qabul qilinishi, ya’ni estetik idrok jarayoni. Shunga asosan, adabiy faoliyatni shunday munosabatlar zanjiri sifatida tasavvur qilish mumkin, ya’ni: voqelik–yozuvchi–asar–o‘quvchi”[Хованская З.И., 1980. 14]. “Badiiy matnning lingvopoetik tahlilida ana shu to‘rt halqali zanjir hamisha e’tiborda bo‘lmog‘i lozim”[Йўлдошев М., 2009. 164].

Lingvopoetikaning konseptual apparati ushbu yo‘nalishning o‘ziga xos xususiyatlarini aks ettirish imkonini beradi. Lingvistik stilistika bilan solishtirganda tadqiqot adabiyot va lingvopoetik tahlilning mavjud usullari badiiy matnga xos alohida elementlarga ham, matnni bir butun sifatida ko‘rib chiqishga ham tegishlidir. Yozuvchi o‘zi his qilayotgan tuyg‘ularni o‘quvchiga badiiy, obrazli tarzda yetkazib berish maqsadida, o‘quvchini o‘zining hissiy olamida bimalol sayr qila olishini ta’minlash maqsadida, uni ham xuddi o‘zidek fikrlar ummoni oldida yolg‘iz qoldirib, ulardan turli ma’no injaliklarini tera olishlariga imkon yaratish maqsadida ham o‘zi yaratayotgan badiiy asarida turfa xil ko‘rinishda g‘ayriodatiy birikmalarni yuzaga keltiradi va tilning estetik vazifasini namoyon etishi uchun ajib bir qulayliklar hadya etadi. Bunday birikmalarni tuza olish yozuvchidan o‘ziga xos katta mahorat, so‘zlar bilan so‘zlasha olish, so‘zlarning “tilini topa bilish” (ta’kid bizniki, M.A.) kabi iqtidor, aqliy salohiyat, badiiy ustamonlikni talab etadi. Har qaysi ijodkor ham bu kabi birikmalarni yangidan yarata olmasligi mumkin. Lekin shunday g‘ayriodatiy birikmalar ham bo‘ladiki, ular bir muallif tomonidan yaratilib, keyinchalik boshqa bir yozuvchi tomonidan asarlarida foydalanishi, bora-bora bu birikmalar iste’molga ko‘proq aralashishi

natijasida badiiy nutqdan tilga ko'chib qoladi. Shunday bo'lsa-da, u baribir g'ayriodatiy birikmaligicha qoladi, faqat o'zidagi estetiklik darajasini biroz kamaytirgan bo'ladi. Misol uchun, *shirin so'z, achchiq haqiqat, shirin muomala* kabilar (bu birikmalar ham aslida g'ayriodatiydir, sababi, so'zning, haqiqatning, muomalaning hech ta'mini totib ko'rish mumkinligi haqida o'ylab ko'rganmisiz, bular mavhum tushunchalar bo'lgani holda ularda ta'm bo'lmasligi tabiiy hol; ularda bor-yo'g'i estetiklik pasaygan).

Bizningcha, badiiy matnda g'ayriodatiy birikmalarning vazifasi shundan iboratki, eng avvalo, ularga asarga xos bo'lgan estetiklik yuklanadi, matnning o'quvchiga badiiy ta'sir etish bilan matnning ekspressiv-emotsionalligi hamda lingvopoetik qimmatini ortishiga sababchi bo'ladi. Fikrimizning isboti o'laroq, quyida shunday g'ayriodatiy birikmalar ishtirok etgan badiiy nutq parchalaridan namunalar keltiramiz.

Asadbek darg'azab edi. – Asadbek tahdidning libosini kiyib olgan edi. (T.M., «Shaytanat»). Keltirilgan ikki gapga diqqat qilaylik, birinchi gapda so'zlar o'z ma'nolarida, o'quvchi bir qarashdayoq gapning mazmunini to'liq tushunadigan holda aniq va qisqa gap shaklini ko'ramiz, ya'ni Asadbekning g'azablangani haqidagi axborotni to'g'ridan-to'g'ri qabul qilamiz. Ikkinchi gapda esa, yuzaki o'qilganda, Asadbek qandaydir libos kiyib olgani bilan tanishamiz, yaxshiroq e'tibor bersak, bu libos tahdidning libosi ekanligini bilamiz, ana shu yerda bizning tafakkurimizda uyg'onish sodir bo'ladiki, bunga sabab "tahdidning libosi" birikmasidir. Ana shunda anglash bosqichiga o'tamiz: tahdid – bu "qo'rqitish, do'q-po'pisa" kabi ma'nolarni bildiruvchi mavhum ot sanalgani holda uning qanday libosi bo'lishi mumkin, degan savol tug'iladi; libos so'zi ongimizda "kiyim, ust-bosh" ni anglatuvchi so'z sifatida namoyon bo'ladi. Muallif so'zlardagi o'z va ko'chma ma'nolarini ifodalay olish imkoniyatlarini bilgani holda, libos so'zining badiiy tasvirda o'ragan, qoplagan narsa ma'nosiga nisbatan ko'chma ma'no hosil qilishini asos qilib oladi va uni tahdid so'ziga qo'shish orqali qahramon tashqi qiyofasining badiiy tasvirini berishda o'rinli foydalanadi.

Yana bir asardan keltirilgan parchada mavhum tushuncha asosidagi epitetni ko'ramiz: *Shu qadar og'ir sukunat cho'kadiki, quloqlaringiz shang'illab ketadi* (O'.H., «Dunyoning ishlari»). Bu gapdagi g'ayriodatiy birikma – *og'ir sukunat* birikmasi bo'lib, sukunat so'ziga og'ir so'zi sifatlovchi bo'ladi. Bizga ma'lumki, sukunat so'zi jimlikni, hech qanday ovozsizlikni anglatuvchi mavhum ot sanaladi, uning belgisini bildiruvchi sifatida berilayotgan so'z esa "mashaqqatli, uqubatli, azobli" kabi ko'chma ma'nolarni bera olgani bois sukunat so'ziga biriktirilgan. Gapning so'ngida keltirilgan so'zga e'tibor bersak, bu og'ir sukunat quloqlarni shang'illatib yuborishi mumkinligi aytiladi, bu esa sukunatning haqiqatan ham azobli ekanligini anglatadi va birikma tarkibida og'ir so'zi ayni shu ma'noni ifodalayotganiga ishoradir. Bu gap tarkibida tushunchalar o'rtasida zidlanish munosabati bor, ammo gapda mantiqiy badiiy-estetiklikni ana shu og'ir so'zidagi yashirin semaning faollashishi yuzaga chiqarmoqda.

Muallif gap tuzishni, gapda fikr ifodalashning ajib bir usulini tanlaydiki, pirovardida matnni o'qigan kishi tamomila asarga kirishib ketadi, qahramonning hayoti bilan yashashni, uning ichki kechinmalarini his eta boshlaydi. Badiiy matnning o'quvchini o'ziga ana shunday ipsiz bog'lashi, turfa xil g'ayriodatiy birikmalardan foydalanish orqali matndan yuksak estetik zavq olishiga erishishi, albatta, yozuvchining badiiy mahorati, yuksak badiiy tafakkuri, so'zlarga estetik vazifa yukini yuklay olishi bilan sodir bo'ladi. Bu esa badiiy matn lingvopoetikasining asosiy bosh masalalaridandir.

Foydalanilgan adabiyotlar ro'yxati:

1. Аврорин В.А. Проблемы изучения функциональной стороны языка. – Л., 1975. – С. 34.
2. Gariper C. Nazim Hikmet'in Ortulu Ideolojik Dili / Kiiltir Tarihimizde Gizli Diller ve Sifreler. – Istanbul, 2008. – P. 159.
3. Qarang.: Фирсова Н.М. Введение в грамматическую стилистику современного испанского языка. – М.: Высшая школа, 1981. – С. 8-9.

4. Гореликова М.И., Магомедова Д.М. Лингвистический анализ художественного текста. – М.: Русский язык, 1989. – С. 5.
5. Кожин А.Н. Язык художественной литературы как эстетически стимулируемая форма существования литературного языка / Структура и функционирование поэтического текста. Очерки лингвистической поэтики. – М.: Наука, 1985. – С. 12.
6. Махмудов Н. Нутқ маданияти ва тилнинг эстетик вазифаси // Филология масалалари. – Тошкент, 2006. №2(11). – Б. 47, 48, 51.
7. Хованская З.И. Анализ литературного произведения в современной французской филологии. – М.: Высшая школа, 1980. – Б. 14.
8. Йўлдошев М. Бадиий матнинг лингвопоэтик таъкики: Филол. фан. д-ри ... дисс. – Тошкент, 2009. – Б. 164.
9. Shavkatovna, A. M. (2023). Gʻayriodatiy birikmalarning badiiy matnda lingvopoetik aktuallashuvi. Journal of Integrated Education and Research, 2(3), 129-132.
10. Shavkatovna, A. M. (2023). Morphological Characteristics of Unusual Compounds Found in Literary Text. World of Science: Journal on Modern Research Methodologies, 2(3), 83-88.
11. Shavkatovna, A. M., & Nurjahon, M. (2023). Badiiy matnda uchraydigan gayriodatiy birikmalarning morfologik xususiyatlari. Ustozlar uchun, 15(1), 266-270.
12. Munira Sh. Akbarova, . (2022). Linguopoetic features of unusual compounds in the literary text. current research journal of philological sciences, 3(05), 73–76. <https://doi.org/10.37547/philological-crjps-03-05-17>

LINGVISTIK EKSPERTIZANING KRIMINAL VAZIYATLARGA TATBIQI MASALASI

Musulmonova Kamola

Filologiya fanlari bo'yicha falsafa doktori(PhD)

Alisher Navoiy nomidagi Toshkent davlat

o'zbek tili va adabiyoti universiteti

Annotatsiya: Huquqshunoslik sohasida jinoyatlarga ta'rif berish va tasniflash yuridik aspektida olib boriladi. Jinoyatlar sirasida bevosita og'zaki yoki yozma matnlarning muayyan ko'rinishlari mavjudki, huquqiy baho berishdan avval lingvistik tekshiruv o'tkazish zarurati paydo bo'ladi. Ana shunday ziddiyatli holatlarda lingvistik ekspertiza tayinlash va o'tkazish amaliyoti jahonda va respublikamizda yo'lga qo'yilgan. Biroq aynan qanday kriminal vaziyatlar lingvistik ekspertiza obyekti bo'la olishi masalasi hal etilmagan. Maqolada og'zaki yoki yozma matnlar bilan bog'liq kriminal vaziyatlarda lingvistik ekspertizaning tatbiqi masalasi yoritilgan.

Kalit so'zlar: Lingvistik ekspertiza, sud, huquq, kriminalistika, til jinoyatlari, lingvistik ekspertologiya, grafologiya, konfliktli matn.

Annotation: In the field of law, the definition and classification of crimes is carried out from a legal aspect. Among the crimes, there are certain manifestations of direct oral or written texts, which necessitate a linguistic examination before making a legal assessment. In such conflicting cases, the practice of appointing and conducting linguistic expertise has been established in the world and in our republic. However, the question of what exactly criminal situations can be the object of linguistic expertise has not been resolved. The article discusses the issue of the application of linguistic expertise in criminal situations related to oral or written texts.

Keywords: Linguistic expertise, court, law, criminalistics, language crimes, linguistic expertology, graphology, conflict text.

Аннотация: В области юриспруденции определение и классификация преступлений осуществляется с юридической точки зрения. Существуют определенные формы прямых устных или письменных текстов в преступлениях, которые требуют проведения лингвистической экспертизы перед вынесением юридической оценки. Практика назначения и проведения лингвистической экспертизы в подобных спорных случаях сложилась во всем мире и в нашей республике. Однако вопрос о том, какие именно криминальные ситуации могут быть объектом лингвистической экспертизы, до сих пор не решен. В статье рассматривается применение лингвистической экспертизы в криминалистических ситуациях, связанных с устными или письменными текстами.

Ключевые слова: Лингвистическая экспертиза, суд, право, криминалистика, языковые преступления, лингвистическая экспертология, графология, конфликтный текст.

Lingvistik ekspertologiya amaliy tilshunoslikning alohida bir yo'nalishiga aylanib ulgurdi. Lingvistik ekspertiza tilning ijtimoiy fan sifatida keng qo'llanilishi, shuningdek, fonetika, fonologiya, leksikologiya, morfologiya, sintaksis, semantika, stilistika, pragmatika, psixolingvistika, neyrolingvistika, sotsiolingvistika, dialektologiya va korpus tilshunosligi kabilar bilan tanishishni talab qiladi. Lingvistik ekspertizaning aynan qanday kriminal vaziyatlarda qo'llanilishi masalasi ziddiyatli. Umuman olganda politsiya va gumondorning o'zaro ta'sirini talqin qilish, politsiya hisobotlarini tekshirish va nutq namunalari, mualliflik atributini chuqurroq o'rganishda lingvistik ekspertizaning roli muhim ahamiyat kasb etadi.

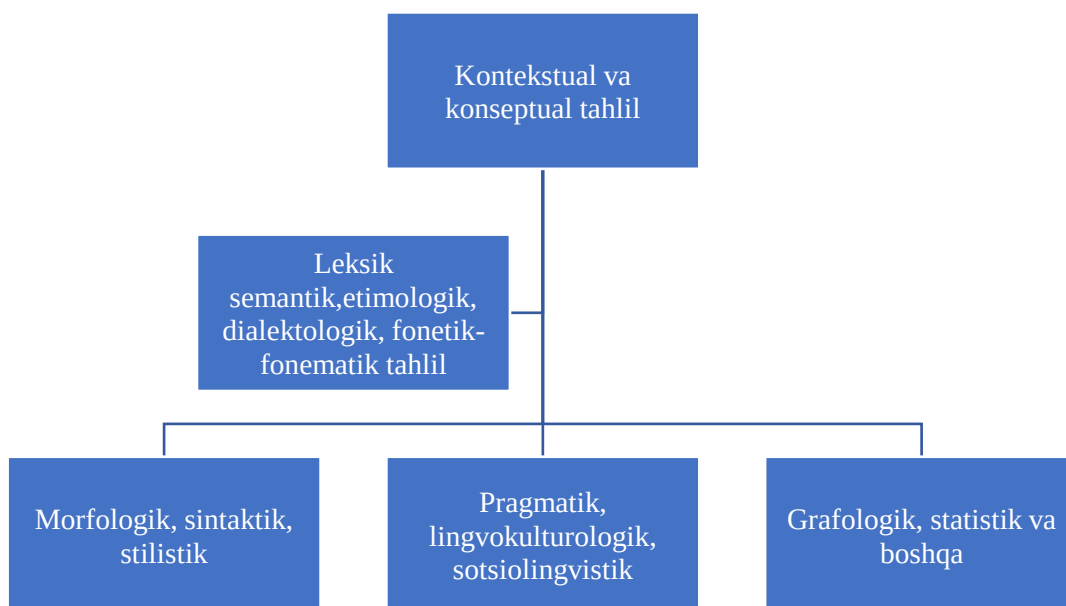
Lingvistik ekspertizaning boshlanish va tugash chegaralarini aniq belgilab bo'lmaydi. Antik davrlarda Qadimgi Yunonistonda dramaturglar bir-birini plagiatlikda ayblashgan. Shuningdek, Shekspir asarlarining muallifi masalasi ham uzoq vaqtlar davomida muhokamalarg

sabab bo'ldi. o'sha davrlarda matnni yuidik va kriminal aspektda tekshirishning yagona tizimli mezoni mavjud emas edi. Turli jinoyatlar, aynan qotillik, suitsid xatlar bilan bog'liq holatlarda tilning tatbiqi masalasi dastavval John Olsson tadqiqotlarida o'z ifodasini topdi.[Olsson John., 2008:206] U lingvistik ekspertizani quyidagicha izohlaydi: “Sud lingvistikasi, to'g'rirog'i, lingvistik bilimlarni ma'lum bir ijtimoiy muhitga, ya'ni huquqiy forumga tatbiq etishdir. Keng ma'noda aytishimiz mumkinki, Sud-tibbiy tilshunoslik til, jinoyat va huquq o'rtasidagi o'zaro bog'liqlik bo'lib, bu yerda qonun huquqni qo'llash, sud ishlari, qonunchilik, nizolar yoki qonun hujjatlarini ko'rib chiqishni o'z ichiga oladi, hatto qonunning ba'zi bir buzilishi yoki huquqiy chora izlash zaruriyati bilan bog'liq bo'lgan nizolar ham lingvistik ekspertiza ko'zgusida tekshiriladi.”

Yan Svartvikning jinoiy kontekstni lingvistik bilimlar orqali jiddiy amaliy tekshiruvlaridan so'ng lingvistik ekspertizaning rivojlanishi sekin kechdi. Turli sud ishlarida lingvistik bilimlarni tatbiq qilish haqida dastavval ayrim mulohazalar bildirildi. [Eades., 1994:26], [Robert Shuy., 1993: 202].

Tilshunoslik borasidagi bilimlarni sud-huquq, kriminalistikaning aynan qanday sohalariga tatbiq etish mumkin degan savol munozarali. Bugungi kunda Respublikamizda faqat haqorat, kamsitish va obro'sizlantirish bilan bog'liq holatlardagina lingvistik ekspertiza o'tkaziladi. Aslida esa to'g'ri metodika tanlash va nazariy bilimlarni amaliyotga to'g'ri tatbiq qilish ko'plab muammolarning obyektiv yechimiga xizmat qiladi. Lingvistik ekspertiza o'tkazish mumkin bo'lgan sud ishlarini belgilashda yuzaga kelishi mumkin bo'lgan muammoli vaziyatlarni belgilash, muammo indikatorini baholash lozim. Bunda og'zaki yozma matnning tuli ko'rinishlari ishtirok etadigan sud, tergov va tergovgacha bo'lgan jarayonlar tizimlashtiriladi, qanday holatlarda lingvistik ekspertiza o'tkazish zarurati borligi analiz qilinadi. O'zbekiston Respublikasi Jinoyat kodeksida og'zaki yozma matn ko'rinishlari ishtirok etadigan yoki ishtirok etishi mumkin bo'lgan jinoyatlar mavjud: 1. Hayotga qarshi jinoyatlar, 2. Oilaga, yoshlarga va axloqqa qarshi jinoyatlar; 3. Shaxsning ozodligi, sha'ni va qadr-qimmatiga qarshi jinoyatlar 4. Fuqarolarning Konstitutsiyaviy huquq va erkinliklariga qarshi jinoyatlar 5. Tinchlik va xavfsizlikka qarshi jinoyatlar.

Yuqorida ko'rsatilgan jinoyatlar mazmunida bevosita matn bilan ish ko'riladi. Ovozli xabarlarlar, chat yozishmalati, og'zaki muloqotlar, ijtimoiy tarmoq matnlari, media makon mahsulotlari va umuman aytganda, og'zaki va yozma muloqotning bsohqa ko'rinishlari ishlatiladigan, yoki ishlatilishi mumkin bo'lgan holatlarda lingvistik ekspertiza o'tkazilishi lozim. Kriminal vaziyatlarda lingvistik ekspertiza o'tkazish usullarini quyidagicha belgilash mumkin:



To'g'ri tanlangan usul va metodika orqali turli jinoyatlarga oid quyidagi muhim dalil vazifasini bajaruvchi ma'lumotlarni olish mumkin:

1. Matnda qo'llangan so'zlarning etimologik, kontekstual ma'nolarini aniqlash orqali oshkora va yashirin(implitsit) mazmunni aniqlash.

2. Matnda qo'llangan leksik birliklarning semantik mundarijasini belgilash orqali umumiy pragmatik mazmunni belgilash.

3. Matnning orfografik xususiyatlarini tekshirish orqali muallif identifikatsiyasini amalga oshirish.

4. Matnda muayyan fonetik, leksik, morfologik, sintaktik birliklarning qo'llanilish chastotasi va shaklini tekshirish, hamda ma'lumotlarni umumlashtirish orqali muallif atributlarini belgilash, anonimlik masalasini oydinlashtirish.

5. Matnni pragmatik, sotsiolingvistik xususiyatlarini inobatga olish, statistik tahlillarni o'tkazish, leksik birliklarning semantik mundarijasini belgilash orqali yashirin kodlarni aniqlash.

6. Matnda lingvistik birliklarning dialektal xususiyatlarini tahlil qilish orqali muallif hududiy xosligi masalasini oydinlashtirish.

Xulosa qilib aytganda, nutqiy xatti-harakatlarning jinoyat sifatida baholanishi so'zlovchining kommunikativ maqsadi va uning ijtimoiy xavfliligi darajasiga ko'ra belgilanadi. Nutqiy mahsulotni tekshirish, lingvistik birliklarning semantik mundarijasini belgilash, hamda kommunikativ maqsadni baholash lingvist-ekspertning vazifasiga kiradi. Uning ijtimoiy xavfliligi va hamda qilmishning jinoiyligini baholash esa masalaning yuridik jihati sanaladi.

Foydalanilgan adabiyotlar:

1. Olsson, J. 2009. Word Crime: Solving Crime through Forensic Linguistics. New York and London: Continuum International Publishing Group. —. 2008. Forensic Linguistics. New York and London: Continuum International Publishing Group.
2. Eades, D. 2008. Courtroom talk and neocolonial control. Berlin and New York: Mouton de Gruyter. —. 2000. I don't think it's an answer to the question: Silencing aboriginal witnesses in court. Language in Society, 2000 (29), pp. 161-195.
3. Shuy, Roger. (1993). Language Crimes: the Use and Abuse of Language Evidence in the courtroom. Cambridge, MA: Blackwell. --- (1997). The Language of Confession, Interrogation, and
4. Deception (vol.2). Sage publication. Shuy, Roger. (2002). Applied Linguistics in the Legal Arena. Incadlin, Chris and Srikant Saraangi (eds.) Handbook of Applied Linguistics Berlin: Mouton de Gruyter. --- (2005). Creating Language Crimes: How Law Enforcement Uses (and Misuses) Language. New York: Oxford U Press.
5. Momeni, Negar (2011). Forensic Linguistics: A Conceptual Frame of Bribery with Linguistic and Legal Features (a Case Study in Iran). International Journal of Criminology and Sociological Theory, Vol.4, No.2, December 2011. 713-744.
6. Peter M. Tiersma, Lawrence M. Solan, The language of crime. eds., The Oxford Handbook of Language and Law, 2012 Brooklyn Law School, Legal Studies Paper No. 263 Loyola-LA Legal Studies Paper No. 2012-14

TEZAURUSLARDA TUR-JINS MUNOSABATI

Xolmurodova Iroda Axtam qizi

O'zbekiston Milliy universiteti

Kompyuter lingvistikasi va

Amaliy tilshunoslik ixtisosligi 1-kurs magistranti

E-mail: ixolmurodova703@gmail.com

Annotatsiya: ushbu maqolada tezaurus atamasining ma'nolari va turlari undagi tur-jins munosabatlari (giponim va giperonim) haqida ma'lumotlar berilgan va shaxs otlari misolida tadqiq etilgan. Bundan tashqari, tezaurlardagi sinonimik va antonimik munosabatlar ham ochib berilgan.

Tayanch so'zlar: giponim, giperonim, giperonimiya, tur-jins, butun-qism, tezaurus, xolonimiya, melonimiya,

Annotation: This article provides information about the meanings and types of thesaurus terms, their species-genus relationships (hyponym and hyperonym), and explores the topic using the example of personal nouns, further clarifying the topic through examples. In addition, synonymous and antonymic relationships in thesauruses are also revealed.

Keywords: hyponym, hypernym, hypernymy, species-genus, whole-part thesaurus, holonymy, melonymy,

Аннотация: В статье представлена информация о значениях и типах терминов тезауруса, их типо-гендерных связях (гипоним и гипероним), а также рассматривается тема на примере личных имен существительных, что дополнительно поясняет тему с помощью примеров. Кроме того, выявляются синонимические и антонимические отношения в тезаурусах.

Ключевые слова: гипоним, гипероним, гиперонимия, вид-род, целое-часть, тезаурус, холонимия, мелонимия

Kirish.

Dunyo tilshunosligida kompyuter lingvistikasi va korpus lingvistikasi muammolarini o'rganish o'tgan asrning 40-yillarida boshlangan. Jumladan, XX asrning 60-yillarida mazkur jarayon jadallashib, XXI asrning boshlarida o'z tarkibida millionlab so'zlarni aks ettiruvchi yuzdan ortiq til korpuslari paydo bo'ldi. Kompyuter lingvistikasi fani globallashuv davri muammolarini hal qilishga yo'naltirilgan, yoshlarning ham texnikani, ham xorijiy tillarni mukammal o'rganishlarini ta'minlaydigan soha hisoblanadi. Kompyuter lingvistikasining asosiy maqsadi lingvistik masalalarni yechishning kompyuter dasturlarini ishlab chiqish hamda bu dasturlar uchun ma'lumotlar bazasini yaratishdan iborat.

Sun'iy intellektda avtomatik tarjima, kompyuter tahlili va uni tahrirlash, tezaurus, elektron lug'atlar singari imkoniyatlari kengaydi. Bilamizki, lingvistik resurslar – lug'atlar, tezaurlar, ontologiyalar shaklida ifodalanadi. Shuningdek, ontologiya, tezaurus, WordNetlar semantik texnologiya instrumentariylari hisoblanadi. [Abduraxmonova, Qodirova, 2024; 30].

“Tezaurus yunoncha “xazina” ma'nosini anglatishi manbalarda qayd etilsa-da, Duglas Xarperning ta'kidlashicha, tezaurus yunoncha fe'l so'z turkumidan kelib chiqqan bo'lib, “joylashtirish, joylashish” ma'nolarini anglatadi. Robert Beekesning keltirishicha, ushbu termin XVI-XIX asrlarda tezaurus lug'at yoki ensklopediya ma'nolarida qo'llanilgan”. “Tezaurus so'zlardagi ichki munosabatlarni o'z ichiga oluvchi boshqariladigan lug'at turi bo'lib, tezaurus tarkibidagi so'zlar alfavit tartibida emas, balki mavzu va mavzular majmui sifatida shakllangan lingvistik lug'atdir. Tezaurus lug'atlarda har bir so'z sinonim, omonim, antonimik tahliliga ega bo'lib, qolaversa, etimologiyasi va so'zlar o'rtasidagi valentlik munosabatlari ham ifoda etiladi”.

“Tezaurus asli til semantikasini belgilaydi. Tezaurusda soʻzlar oʻrtasidagi semantik maʼno jihatidan bogʻlanishlar sinonimlar, antonimlar, giponimlar, giperonimlar, guruhlar orqali farqlanadi” [Abduraxmonova, Qodirova, 2024; 30-31]. Tezaurus soʻzlar va ularning sinonimlari (oʻxshash maʼnoli soʻzlar) va antonimlar (maʼnosi qarama-qarshi boʻlgan soʻzlar)ning tizimli tashkil etilishini taʼminlovchi mos yozuvlar vositasi yoki resursidir. U yozuvchilar, til oʻrganuvchilar va muqobil soʻz va iboralarni taklif qilish orqali oʻz soʻz boyligini oshirishga intilayotgan har bir kishiga yordam berish uchun moʻljallangan. [Abdurakhmonova, Alisher and Toirova, 2022; 87]. Tezaurus odatda soʻzlarni kontseptual toifalar boʻyicha tartibga soladi va foydalanuvchilarga tegishli atamalarni boshqarishga va ular orasidagi aloqalarni aniqlashga yordam beradigan ierarxik tuzilmani taqdim etadi. Misol uchun, “baxt” soʻzi “quvonch”, “sevgi” va “muhabbat” kabi sinonimlar bilan “hissiyotlar” turkumiga kiritilishi mumkin. Tezaurusning asosiy eʼtibori sinonimlarga qaratilgan boʻlsada, u foydalanuvchining asl soʻz haqidagi tushunchasini kengaytiradigan taʼriflar, foydalanish misollari va tegishli soʻzlar kabi qoʻshimcha maʼlumotlarni ham berishi mumkin. Tezaurus turli formatlarda, jumladan bosma kitoblar, onlayn maʼlumotlar bazalari va integratsiyalashgan dasturiy vositalarda mavjud. Ular umumiy yoki ixtisoslashgan boʻlishi mumkin, ular fan, huquq yoki tibbiyot kabi muayyan sohalarga qaratilgan. Tezaurusning baʼzi mashhur misollari orasida Roget tezaurusi, Wordnet tezaurusi, ingliz tilining Oksford tezaurusi va Merriam-Vebster tezauruslari mavjud. Tezauruslarning bir nechta turlari mavjud boʻlib, ularning har biri oʻzining dizayni va moʻljallangan auditoriyasiga qarab muayyan maqsadlarga xizmat qiladi. Bu yerda tezaurusning bir nechta keng tarqalgan turlari mavjud:

1. Umumiy tezaurus: tezaurusning bu turi keng lugʻatni qamrab oladi va umumiy tildan foydalanish uchun moʻljallangan. U bir nechta domenlardagi soʻzlarning sinonimlari va antonimlarini taqdim etadi va kundalik tilga murojaat qilish va lugʻatni kengaytirish uchun moʻljallangan.

2. Mavzuga xos tezaurus: Ushbu turkumdagi tezauruslar tibbiyot, huquq yoki texnologiya kabi maʼlum bir mavzu yoki sohaga qaratilgan. Ular qiziqish sohasiga xos ixtisoslashgan terminologiya, sinonimlar va tegishli atamalarni taqdim etadi.

Ushbu tezauruslar sohaga oid lugʻat va aniq atamalarni taklif qilish orqali oʻz sohalarida mutaxassislar, tadqiqotchilar va talabalarga yordam berish uchun moʻljallangan.

3. Iyerarxik tezaurus: tezaurusning bu turi soʻzlarni ierarxik tuzilishda tartibga soladi, bu yerda kengroq terminlar (giperonimlar) va torroq terminlar (giponimlar) olinadi. Bundan tashqari, meronim (qism orqali butun tushuniladi) holonim (butun qismi orqali qism tushuniladi) partonim (qism) kabi iyerarxik tushunchalar atamalar orasidagi iyerarxik munosabatlarni koʻrsatib, foydalanuvchilarga tushunchalar boʻylab harakatlanish va tilning iyerarxik tuzilishini tushunish imkonini beradi.

4. Assotsiativ tezaurus: Assotsiativ tezaurus qattiq iyerarxiyadan koʻra soʻzlar orasidagi assotsiativ munosabatlarga urgʻu beradi. Ular soʻzlarni semantik assotsiatsiyalar asosida birlashtiradi, masalan, odatda birga topiladigan yoki oʻxshash kontekstlarni oʻrtoqlashadigan soʻzlar. Ushbu tezauruslar foydalanuvchilarga iyerarxik jihatdan bogʻliq boʻlmasa ham, kontseptual aloqaga ega boʻlgan soʻzlarni oʻrganishga yordam beradi.

5. Onlayn tezaurus: Raqamli texnologiyaning paydo boʻlishi bilan tezauruslar endi onlayn formatlarda mavjud. Onlayn tezauruslar soʻzlar, sinonimlar va tegishli atamalarning keng maʼlumotlar bazalariga dinamik kirishni taʼminlaydi. Ular koʻpincha qidiruv imkoniyatlari, xatchoʻplar va oʻzaro havolalar kabi qoʻshimcha funksiyalarni oʻz ichiga oladi, bu ularni til maʼlumoti uchun qulay va foydalanuvchilarga qulay manbalar qiladi. Lingvistik tezauruslar va ularning turlarini bilib olgan ekanmiz, endi tezaurus tarixiga nazar tashlab oʻtsak maqsadga muvofiq ish boʻladi [Abdurakhmonova, Alisher and Toirova, 2022; 90-91].

Hozir aynan biz tahlil qilishimiz kerak boʻlgan tezaurus turi bu - iyerarxik tezaurusdir, negaki ushbu turdagi tezauruslarda tur-jins munosabati aks etadi. Avvalambor, tur va jins munosabatlarini tashkil etuvchi semantik birliklar nimalar iborat ekanligini aniqlab oladigan boʻlsak ular quyidagilarni tashkil etadi:

O'zbek tilidagi tezauruslarni tuzishda sinonim, antonim birliklar bazasi uchun yetarli tadqiqotlar, manbalar, lug'atlar mavjud. Lekin giponim, giperonim, xolonim, meronimlar bo'yicha tadqiqotlar mavjud bo'lsa-da, lug'atlar yaratilmagan. Jahon tilshunosligida J.Layonz so'zlararo ma'no munosabatlari doirasida yangi – tur-jins munosabatini ifodalovchi giponimiya hamda giponimik munosabatlar bilan bog'langan so'zlar sirasi, giponimik lug'aviy ma'no guruh(LMG)lar, giponim, giperonim kabi tushuncha va atamalarini olib kirdi [Layonz, 1978; 65]. O'zbek tilshunosligida giponimiya hodisasining o'rganilishi R.Safarova tadqiqotlari bilan bog'liq [Safarova, 1987; 101].

Dunyodagi boshqa tillar kabi o'zbek tilining ham jins-tur munosabatini ifodalash asosida bog'langan lug'aviy birliklari mavjud: ularning so'z birikmasi vositasida ifodalash imkoniyati cheklanmagan. Jins-tur munosabatini ifodalashning leksik imkoniyatlari o'zbek tilida ma'lum lug'aviy birliklar yordamida ifodalanadi. Tilning leksik-semantik tizimida jins belgisining nomini hamda tur belgisining nomini ifodalab kelgan mustaqil lug'aviy birlik mazkur til qurilishidan mustaqil birlik sifatida o'rin oladi. Bunday lug'aviy birlik alohida o'zak morfemaga ega, u tilning mustaqil birligi. Shuning uchun ham bunday lug'aviy birliklarning miqdori har bir tilda qat'iy ravishda belgilab qo'yilgan So'zning giponimik qatori giperonim hamda giponimdan iborat.

Giperonim – jins belgisini bildirgan predmetning nomini ifodalovchi bir necha ma'noni semantik jihatdan umumlashtiruvchi mikrotizimning markaziy so'zi, dominantasi (daraxt).

Giponim – ma'lum jins turining nomini ifodalovchi hamda o'zining semantik tarkibida jins ma'nosini ifodalovchi so'zni birliktirib kelgan, semantik jihatdan giperonimga nisbatan boy bo'lgan lug'aviy birlik (meva, ildiz, shox...) [Safarova, 1987; 103].

Tilshunoslikda butun-qism tushunchasi ham bor butun va qism o'rtasidagi munosabat sistem leksikologiyada meronimiya termini bilan ataladi. Masalan, kompyuter – butun; monitor, klaviatura, sichqoncha, protsessor kabilar – qismdir. Butun uchun xolonim, qism uchun meronim atamaları ishlatiladi. Demak, motor – avtomobil uchun meronim, avtomobilning o'zi esa motor uchun xolonim sanaladi [Abduraxmonova, Qodirova, 2024; 36].

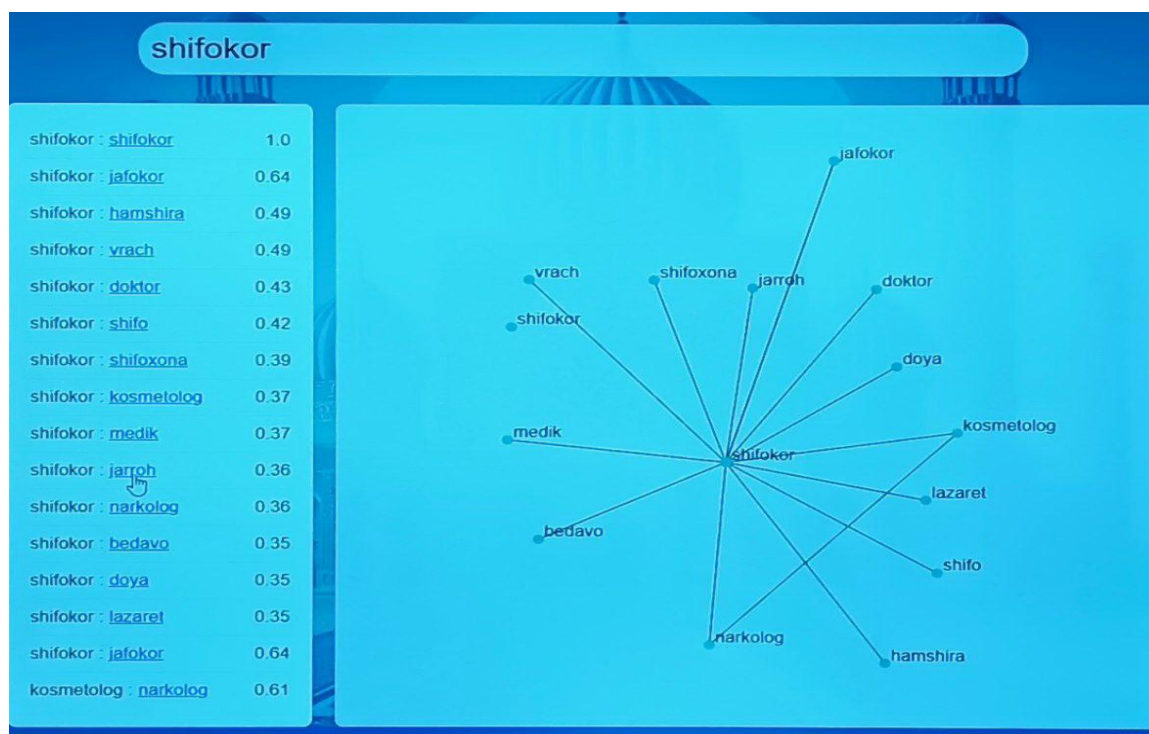
Tur-jins bilan qism-bo'lakni bir biridan farqlashimiz lozim, ikkala munosabat ham ikki xil vazifani bajaradi, ammo tezaurus tuzishda ikkalasidan ham vazifasiga ko'ra foydalaniladi. Quyida shaxs otlari misolida giponim va giperonimlarini tahlil etamiz.

T/R	Giponim	Giperonim
1	ota-ona, aka-uka, farzand	oila
2	o'qituvchi, shifokor, arxetektor	shaxs otlari/kasb
3	erkak- ayol	jins
4	chaqaloq, o'smir	yosh/davr
5	o'zbek, tojik, rus, xitoy	millat
6	boy, kambag'al, o'rtahol	Ijtimoiy mavqe/maqom

Tezaurus termini haqida yana quyidagicha ma'lumotlarni aytish o'rinli. Muayyan tildagi barcha so'zlarni qamrab oladigan, ularning matnda qo'llanish holatlarini to'la to'kis aks ettiradigan lug'at. Jonli tilning lug'at tarkibi doimo o'zgarib turishi sababli Tezaurus faqat o'lik tillar materiallari asosida yoki oldindan belgilab olingan yozma yodgorliklar uchun tuzilishi mumkin. Muayyan yozuv yodgorliklaridan leksikani yoppasiga terib olishga asoslangan lug'atlar (mas., yunon, lotin tillariga tuzilgan lug'atlar) shunday lug'atlardan hisoblanadi. Turkiyshunoslikda A. K. Borovkov, E. I. Fozilov, A. Najiblarning 13—14-asrlar turkiy yozma yodgorliklari asosida tuzilgan lug'atlari ham glossariy, ham Tezaurus belgilariga ega. So'z tanlash prinsipiga ko'ra yozuvchilar yoki ular ijodiga mansub ayrim asar tili bo'yicha tuzilgan lug'atlar ham Tezaurus hisoblanadi; biror bir ilm fan sohasiga oid lug'aviy birlik (so'z) lar mavzu tartibida joylashtirilgan va ana shu birliklar o'rtasidagi semantik (tur-jins, sinonimik va b.) munosabatlar aks ettirilgan ideografik lug'at. Bunda so'zlar alifbo tartibida emas, balki bir mavzu, bir tushunchaga aloqador so'zlarning barchasi bir joyda beriladi: kerakli so'zlar tushunchaga qarab

qidiriladi. Nazariy jihatdan Tezaurus leksik semantik tizimning ehtimoliy modellaridan biridir. Amalda undan yozuvchi individual lug'atini boyitish vositasi sifatida foydalanishi mumkin.

Quyidagi diagrammada "shifokor" so'zi bilan semantik bog'liq bo'lgan so'zlar va ularning ma'nosi o'rtasidagi masofalar o'lchanib ko'rsatilganiga guvoh bo'lishimiz mumkin.



Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Qodirova Z. Ontologiyalar va semantik texnologiyalar (o'quv qo'llanma) Tosh. 2024
2. Abdurakhmonova N. I. Alisher and G. Toirova, Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing, 2022 7th International Conference on Computer Science and Engineering (UBMK), Diyarbakir, Turkey, 2022, pp. 73-75, doi: 10.1109/UBMK55850.2022.9919521.
3. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of Uzbek words. Applied Sciences, 11(19), 9117.
4. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
5. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
6. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus. uz. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.
7. Лайонз Дж. Введение в теоретическую лингвистику. – Москва, 1978.
8. Орифжонова Ш. Ўзбек тилида луғавий градуонимия: филол. фан. номз. дисс. автореф. – Тошкент, 1996. – 21 б.
9. Қиличев Б. Партонимиянинг синонимия, гипонимия ва градуонимияга муносабати // Ўзбек тили ва адабиёти, 2002, 5-сон.
10. Қиличев Б. Ўзбек тилида партонимия: филол. фан. номз. дисс. автореф. – Тошкент, 1996. – 25 б.

11. Сафарова Р. Ўзбек тилида гипонимия: филол. фан. номз. диссер. автореф. – Тошкент., 1987. – 26 б.
12. “WordNet Statistics”. Wordnet.princeton.edu. Retrieved 2018-06-22.

DOSTONLAR LEKSIKASIDA UCHRAYDIGAN O'XSHATISHLI BIRLIKLARNING TIPLARI

Xudayberganova Dildora
O'zMU dotsent v.b., PhD.

Annotatsiya: Maqolada Xorazm dostonlari tilida uchraydigan turli poetik maqsadlarga xizmat qiluvchi tasviriy ifodalar, muayyan qolipga tushib qolgan birikma shaklidagi diskurs vositalarining o'ziga xos lingvistik jihatlari haqida mushohada yuritiladi. Mazkur masalani yoritishda qo'lyozma va toshbosma dostonlar matni materiallari asos qilib olingan.

Kalit so'zlar: Xorazm, doston, o'xshatishli birlik, sodda tarkibli o'xshatishli birlik, murakkab tarkibli o'xshatishli birlik, qiyosiy tahlil, dialektizm, etnografizm.

Abstract: The article examines the specific linguistic aspects of figurative expressions serving various poetic purposes and compound discourse devices that fall into a certain pattern in the language of Khorezm epics. In covering this issue, materials from manuscript and lithographed epic texts were used as a basis.

Keywords: Khorezm, epic, simile, simile with simple content, simile with complex content, comparative analysis, dialectism, ethnography.

Абстракт: В статье рассматриваются специфические языковые аспекты образных выражений, служащих различным поэтическим целям, а также сложные дискурсивные приемы, которые в языке хорезмийских эпосов подчиняются определенной закономерности. При освещении этого вопроса за основу были взяты материалы рукописных и литографированных эпических текстов.

Ключевые слова: Хорезм, эпос, сравнение, сравнение с простым содержанием, сравнение со сложным содержанием, сравнительный анализ, диалектизм, этнография.

Hozirgi kunda folklor asarlari, xususan, dostonlar tilini tekshirish orqali jonli o'zbek tilining eng noyob birliklarini, unutilib ketayotgan turli funksiyadagi lisoniy vositalarni topib tekshirish, tilimizning jozibasini yana bir bor namoyon qilish imkoniyati yuzaga kelmoqda.

Badiiy matnlar hajm belgisiga ko'ra ikki turga ajratiladi: kichik hajmli minimal matnlar (*mikromatn*) va hajman katta bo'lgan maksimal matnlar (*makromatn*). Dostonlar matnlari nafaqat hajm jihatdan, balki ichki mazmun, xarakterlar tasviri, ruhiyat talqini jihatidan ancha mukammaldir⁸². Badiiy asarlar matnlari qanday tuzilishidan qat'i nazar, kitobxonga ta'sir etishga mo'ljallangan bo'ladi va shunga mos ravishda lingvistik vositalar tanlanadi⁸³.

Odatda, o'xshatishli tasvirlarni yuzaga keltirishda ko'plagan lingvopoetik priyomlar mavjud. Bunday o'xshatishli tasvirlar, ko'pincha, qaysidir personajning shakli-shamoyili, mahorati, hunari, bahodirligi yoki go'zalligini asoslashga qaratilgan bo'ladi. Ushbu xususiyatlarni tasvirlashga moslashgan lingvistik vositalar o'z xarakteriga ko'ra jins nuqtayi nazaridan farqlanadi. Agar ma'shuqaning go'zalligi tasvirlari keltirilayotgan bo'lsa, "*olma kabi*", "*oydek*", "*tovusdek*" kabi lingvistik vositalar "yuz" tasviri uchun qo'llaniladi. Hech qachon oshiqning yuzi olmaga yoki oyga o'xshatilmydi.

Shuningdek, *sherdek*, *arslondek*, *ajdahodek*, *minordek* kabi lisoniy tasvirlash vositalari, albatta, erkak qahramonlar tasviri uchun qo'llaniladi. Shularni inobatga olgan holda aytish mumkinki, dostonlarimiz matnlarida uchraydigan arxaik ko'rinishdagi o'xshatishli tasvir vositalari o'ziga xos gender xususiyatlarini namoyon qiladi.

Dostonlarimiz tilida uchraydigan o'xshatishli tasvirlar alohida qimmatga ega. Ularning ko'pchiligi *kabi* yordamchisining arxaik shakllari *kibi*, *misli*, *yanglig'*, shuningdek, ko'pgina

⁸² Йўлдошев М. Бадий матн ва унинг лингвопоэтик таҳлили асослари. – Тошкент: Фан, 2007. – Б.12.

⁸³ Хамидов А. Мурод Муҳаммад Дўстнинг Лолазор романи лингвопоэтикаси. – Т.: Наврўз, 2019. – Б.11.

turkiy tillar uchun umumiy bo'lgan **kimin** yordamchilari ishtirokida, ba'zan esa **-dek**, **-day** affiksining arxaiklashgan shakli **-dayin** yordamida yuzaga keltiriladi. Xorazm dostonlari tilidagi ayrim o'rinlarda **-dayin** affiksining dialektal ma'nodoshi (varianti) **-din** (Xorazm shevalarida o'xshatish ma'nosini hosil qiluvchi) yordamida yuzaga keltiriladi.

Dostonlar tilidan o'rin olgan badiiy o'xshatishlar yuzaga keltirilgan lisoniy poetik birliklar orasida eng rang-barang shakllar, asosan, yorning (ma'shuqaning) yuzi (jamoli), qaddi-qomati, qadam tashlashi, kuylashi kabilarni tasvirlashda qo'llaniladi. To'plangan misollar orasida dastavval e'tiborni tortgani ma'shuqaning raftori tasviriga moslashgan murakkab qurilma shaklidagi lisoniy vosita bo'ladi. Quyidagi satrlar "Xirmondali" dostonidan olingan:

O'n to'rt gejal oy kabi, Ot-yarog'i ham shay kabi,

Husni davlatli boy kabi, Gazam har yona-har yona...(284-bet).

Ushbu parchada **kabi** ko'makchisi yordamida hosil qilingan bir necha birliklar bor, ammo konkret yorning jamoli tasviriga aloqador dialektal xususiyatlarni aks ettirgani "*o'n to'rt gejal oy kabi*" birikmali ifoda sanaladi. Yuqoridagi birikmali ifodaning tarkibi to'liq turkiy so'zlardan tashkil topganligi alohida qimmatga ega. Biroq bu degani boshqa tillarga aloqadorlari uchramaydi degani emas. Xorazm dostonlari tilida o'zga tillarga aloqador yuz tasvirlari birikmali tarzda ishlatilgan o'rinlar ham ko'p.

"Oshiqnoma"ning 1-kitobidan o'rin olgan "Roiy Chin" dostonidagi parchadagi tasvirlovchi lisoniy vositalarga e'tibor qaratamiz: "...*go'yoki xurshidi xovardek kokillar chin-chin, labi shakar huru g'ilmondek...*" (160-bet). Yuz tasviriga yo'naltirilgan "xurshidi xovar" birikmali ifoda eroniy tillarga aloqador birliklar asosida shakllangan. Bu o'rinda yana bir alohida hodisa kuzatilmoqda. O'xshatish hosil qiluvchi ikkita element bitta vositaga birikib kelgan. Ular: "*go'yoki*" va "*-dek*" elementlaridir. Ulardan biri ortiqcha ishlatilgan, biroq tasvirni kuchaytirish uchun baxshilar bunday hodisalarga (pleonazmgga) keng o'rin berganlar.

Lisoniy poetik tasvir vositalarining yana bir toifasi **kimin** arxaik yordamchisi ko'magida hosil qilinadi. Tanlangan misollarimiz ma'shuqaning faqat yuzi tasvirlarini emas, uning qaddi-qomati chizgilarini ham aks ettiradi. Masalan:

Maral kimin na do'lanib boqarsan, So'zla qurbon o'lam shirin tillara... ("Oshiqnoma" 1-kitob, 56-bet). Yoki:

So'na kimin dol gardaning qoldirma, Har yetar-yetmasa siring bildirma... ("Oshiqnoma" 1-kitob, 79-bet). Ushbu o'rinlardagi ma'shuqaning qomatiga aloqador tasvir vositalari zoonimlar bilan bog'liq lisoniy birliklar ishtirokida shakllangan. Dostonlar leksikasiga oid matnlarda ko'plagan bu kabu tasvir vositalarini uchratish mumkin. Masalan: *tovus*, *xumoyun qushi*, *qumri*, *bulbul*, *to'ti*, *jayron*, *kiyik* kabi zoonimlar ma'shuqaning tasvirlariga xos lisoniy vositalar bo'lsa, *sher*, *arslon*, *lochin*, *yaralangan qoplon*, *ajdaho*, *nar*, *fil*, *bo'ri* kabilar oshiq obrazi tasvirlari uchun qo'llaniladigan lisoniy birliklar sanaladi.

Ma'lumki, xalq dostonlari matnlarida lirik va epik o'rinlar baravar qo'llanadi. Shuning uchun ham ular liroepik asarlar sanaladi. Lirik o'rinlardagi matnlar ko'pincha, personajlar tilidan kuylanadi. Ularning qo'shiq qilib kuylanishi nafaqat doston voqealarida, balki baxshilar ijrosida ham real holatda ro'y beradi. Qo'shiq aytish, kuylash holatlari, asosan, *bulbul* sayrashiga qiyoslanadi, shuning uchun ham *bulbul* leksemasidan ana shu holatlar tasvirida foydalaniladi. "Oshiq Alband" dostonidagi quyidagi parchada *-dayin* arxaik affiksi bilan qo'llanilgan:

Dilbar hayron etding mani husninga,

Bulbuldayin sayrab qo'nsam gulinga... ("Oshiqnoma" 2-kitob, 177-bet), shuningdek, "Xiromon pari" dostonida **kibi** arxaik ko'makchisi bilan qo'llanilganligini ko'ramiz:

Mutribo makong'a bulbul kibi jon sarf etako'r,

Bog'i ishqimg'a mani bir guli ra'no kelajak... ("Oshiqnoma" 2-kitob, 317-bet).

Nofaol qatlamga xos o'xshatishli poetik tasvir vositalari erkak kishilarning tasvirlarida ham faol qo'llaniladi. Biz ushbu o'rinlarda zoonimlar bilan bog'liq lisoniy vositalarning qo'llanishi xususida fikr yuritayotgan ekanmiz, oshiq tasvirlaridan ham ayrim o'rinlarni tilga olib o'tamiz.

“Arab Rayhon” dostonida Go‘ro‘g‘lining tilidan aytilgan misralarda yigitning botirligi *nardek* o‘xshatish vositasi bilan tasvirlanadi:

Ko‘nglim shoddur bugun, Gi‘rot ostimda,

Nardek davusharman maydon ichinda... (Go‘ro‘g‘li, 79-bet). Shu turkumdagi “Bozirgon” dostonida “yarali arslon” birikmasidan foydalaniladi:

Qayg‘u galmas g‘och yigitni ko‘nglina,

Yarali qoplondek ova duch bo‘lar... (Go‘ro‘g‘li, 391-bet) va boshqalar.

Dostonlar matnida uchraydigan o‘xshatishga xizmat qiluvchi lisoniy birliklar orasida personajlar tasvirlarida ularni mashhur asarlar qahramonlari, tarixda yashab o‘tgan real shaxslar yoki to‘qima, afsonaviy timsollarga o‘xshatish hodisalari keng kuzatiladi. Bu hoaltda ham muayyan tarzda jins ko‘rsatkichi farqlanadi. Ma‘shuqa obrazi tasvirlarida *Layli, Shirin, Zulayho, pari* kabi lisoniy vositalardan foydalanilsa, erkaklar bilan bog‘liq o‘xshatishlarda *Yusuf, Ibrohim* va boshqa payg‘ambarlarga, *Majnun, Farhod* kabi asar qahramonlariga, *afsonaviy devlarga* qiyoslovchi lisoniy poetik vositalarga murojaat qilinganligini ko‘ramiz.

Yuqorida tilga olingan tasvir vositalarining xilma-xil shakllari mavjud, ammo biz ishimiz obyektidan kelib chiqib, nafaol birliklar ishtirok etgan ko‘rinishlargagina to‘xtalamiz. Jumladan, *-dayin* ana shunday xarakterdagi arxaik grammatik vositadir. “Sayod va Hamro” dostonida personaj mumtoz adabiyotimizda faol qo‘llanilgan *Majnun* timsoliga o‘xshatiladi:

O‘tgan kunlarima daftar yozaram,

Majnundayin cho‘ldin-cho‘lga kezaram... (“Oshiqnoma” 1-kitob, 14-bet) yoki:

Jamolin barmanam yoxti jahona,

Yusufdayin solma yana zindona... (“Oshiqnoma” 1-kitob, 35-bet).

Ushbu o‘rinlardagi o‘xshatishlar qahramonlarning ko‘rinishi, shakl-shamoyili kabilarni emas, o‘sha timsollarning hayoti, boshidan kechirgan hodisalarga nisbatan ishlatilgan.

Dostonlarimiz matnlarida bunday tipdagi o‘xshatishlarni yuzaga keltiruvchi vosita sifatida *kibi* arxaik yordamchisi qo‘llanilgan holatlar ham mavjud. Yuqorida tilga olgan doston matnida *Layli kibi* shaklida o‘xshatishli lisoniy vosita uchraydi:

Layli kibi Majnunini yitirib,

Qodir mavlon har maqsada yetirib... (“Oshiqnoma” 1-kitob, 78-bet).

Dostonlarimiz tilida arxaik shaklga ega *yanglig‘* yordamchisi ham o‘xshatishli tasvir vositasini shakllantirishda ishlatiladi. Bu hodisa eski o‘zbek tili qonuniyatlari bilan umumiylik xususiyatlariga ega. “Oshiq Albani” dostonidagi quyidagi parchada ana shu holat uchraydi:

*Ko‘rsatib o‘zni **pari yanglig‘** bo‘lib ko‘zdin pinhon...* (Oshiqnoma 2-kitob, 155-bet).

Ba‘zida bir xil o‘xshatishli ifodaning ma‘nodosh shakllarini yonma-yan qo‘llash holatini kuzatamiz. O‘sha kitobdan o‘rin olgan “*Xiromon pari*” dostonidagi quyidagi parchada shunday holat mavjud:

Hur yanglig‘ bir pari monand Ka‘nondin nishon...(320-bet).

Ushbu o‘rindagi o‘xshatishli birliklar tasvirni kuchaytirishga xizmat qilmoqda.

Yuqoridagi fikrlarni umumlashtirgan holda aytadigan bo‘lsak, Xorazm dostonlari tilida turli jihatdan qo‘llanishi nafaol holatga kelib qolgan birliklar mavjud. Ular xoh davr nuqtayi nazaridan, xoh hududiy jihatdan nafaol birliklar bo‘lishidan qat‘i nazar, ular dostonlar tilining serjilo ohanglarini yuzaga keltirishda muhim vazifalarni bajaradi. Ularni to‘plab, turli jihatlarini tekshirish lingvistik tadqiqotlarga qanchalik muhim ma‘lumotlarni bersa, adabiyotshunoslik va poetika sohalariga ham shunday qimmatli materiallar taqdim etadi.

Umumlashtirgan holda aytish mumkinki, dostonlar tilida uchraydigan o‘xshatishni yuzaga keltiruvchi lisoniy vositalarni turli tomondan tekshirish nafaol qatlam birliklari, xususan, dialektal lisoniy birliklarning funksional jihatlarini yana ham kengroq oydinlashtirish imkonini beradi.

Foydalanilgan adabiyotlar:

1. Йўлдошев М. Бадий матн ва унинг лингвопоэтик таҳлили асослари. — Тошкент: Фан, 2007. — Б.12.

2. Ҳамидов А. Мурод Муҳаммад Дўстнинг Лолазор романи лингвопоэтикаси. – Т.: Наврўз, 2019. – Б.11.
3. Ошиқнома. 1 китоб. Нашрга тайёрловчилар: Рўзимбоев С.Р., Эшчонова Г.О., Рўзимбоев С.С., Аҳмедов А. – Урганч: Хоразм нашриёти. 2006
4. Ошиқнома. 2 китоб. Нашрга тайёрловчилар: Рўзимбоев С.Р., Эшчонова Г.О., Рўзимбоев С.С. – Урганч: Хоразм нашриёти, 2006
5. Гўрўғли. Нашрга тайёрловчилар: Рўзимбоев С.Р., Рўзимбоев Х.С., Эшжонова Г.О. – Урганч: Хоразм нашриёти, 2004. – 480 б.
6. Xudayberganova D. Badiiy matnda kirish va kiritmalarning qo'llanishi (Tohir Malikning "Shaytanat" asari misolida). Tamaddun nuri/the light of civilization issn 2181-8258. 2024-yil, 8-son (59). 159-162.
7. Xudayberganova D. Lingvokulturologiyaning nazariy asoslari xususida mulohazalar. Miasto Przyszłości. – Kielce, 2023. 660-663.
8. Xudayberganova D. Linguopoetic studies in the example of the epic "Ravshan". Spectrum Journal of Innovation, Reforms and Development. 2023. 293-297.
9. Xudayberganova D. The linguistic poetics in the epic of Koroghlu. International multidisciplinary journal for research & development. 2023. 585-590.

O'ZBEK VA NEMIS TILLARIDA FRAZEOLOGIK BIRLIKLARNING LINGVOKULTUROLOGIK XUSUSIYATLARI

Abdullayeva Shahlo

O'zbekiston Milliy universiteti
Jurnalistika va o'zbek filologiyasi fakulteti
1-bosqich talabasi

Annotatsiya: Ushbu maqola nemis va o'zbek tillarida frazeologik iboralarining qo'llanilishi haqida umumiy ma'lumotlar beradi. Bu tildagi iboralar nutqning aniqroq va kuchli bo'lishga yordam beradi. Tadqiqot natijasida ushbu tillardagi frazeologizmlarning o'ziga xos jihatlari aniqlanib, ularning lingvokulturologik ahamiyati yoritiladi.

Kalit so'zlar: Nemis tilidagi frazeologik iboralar, o'zbek tilidagi frazeologik birliklar, o'zbek va nemis tilidagi iboralarining o'xshash va farqli jihatlari, qiyosiy tahlil, lingvokulturologiya,

Annotation: This article provides general information about the usage of phraseological expressions in the German and Uzbek languages. These expressions help make speech more precise and expressive. As a result of the research, the unique features of phraseological units in these languages are identified, and their linguistic and cultural significance is highlighted.

Keywords: Phraseological expressions in the German language, phraseological units in the Uzbek language, similarities and differences of phraseological expressions in Uzbek and German, comparative analysis, linguistic and cultural studies.

Аннотация: Данная статья предоставляет общую информацию об использовании фразеологических оборотов в немецком и узбекском языках. Эти выражения помогают сделать речь более точной и выразительной. В ходе исследования выявлены особенности фразеологизмов в данных языках и раскрыто их лингвокультурологическое значение.

Ключевые слова: фразеологические обороты в немецком языке, фразеологические единицы в узбекском языке, сходства и различия фразеологизмов в узбекском и немецком языках, сравнительный анализ, лингвокультурология.

Frazeologiya so'zi yunoncha "phraseo" — nutq, "logos" — ta'limot so'zlaridan olingan bo'lib, frazeologik birliklar haqidagi ta'limotni anglatadi. O'zbek tilida frazeologik birliklar, ya'ni iboralarining qo'llanilishi adabiy tilimizning boy samarasidir. Ulardan foydalanish orqali tilimizning go'zalligi, obrazlilik namoyon bo'ladi.

O'zbek tilida frazeologik iboralarining rivojlanishi va qo'llanilishi haqida bir qator olimlar o'z ilmiy ishlarida muhim ma'lumotlar berishgan⁸⁴. Ular quyidagilar:

Ayniqsa, Sh.Rahmatullayevning "O'zbek tilining frazeologik lug'ati", N.Mahmudovning "O'zbek tilining frazeologik lug'ati" kabi asarlari o'zbek frazeologiyasini o'rganishda asosiy manba hisoblanadi.

Nemis tili frazeologizmiga oid olimlar olimlar tadqiqot ishlar qilgan. Xususan, Sh. S.Imyaminova "Nemis tili frazeologiyasi", SH.Do'sbekov "Nemis so'zlashuv nutqida qiyosiy frazeologik birliklarning ishlatilishi", M.Abdulazizova "Nemis tilidagi frazeologizmlarning tarjima usullari" kabi asarlari nemis tilidagi frazeologizmlarni o'rganishda muhim manba bo'lib, tarjima jarayonida duch kelinadigan iyinchiliklarni yengishda yordam beradi.

⁸⁴ N.G'ulomov "O'zbek tilining frazeologik lug'ati"; A.Madvaliyev, N.G'ulomov. "O'zbek tilining frazeologizmlari"; A.Mamatov "Frazeologiya va uning lingvistik xususiyatlar"; B.Yo'ldoshev "O'zbek tilida frazeologik birliklarning semantik tuzilishi"

Frazeologik iboralarning har bir til leksikasini boyitishda tilning obrazlilikini ta'minlashda muhim ahamiyati bor. Nemis va o'zbek tillari frazeologik fondining shakllanishida ularning tarixi, madaniyati, turmush tarzi va e'tiqodi muhim omil bo'lgan. "Qo'lini ko'tarish" erkin birikmasi bizning tilimizda yordam so'ramoq ma'nosida ishlatilsa, "Die Hand heben" (qo'lni ko'tarish) turg'un birkmasi, odatda, biror narsani so'rash uchun foydalaniladi. Ba'zi frazeologik birliklar o'zbek va nemis tillarida bir xil ma'noga ega bo'lib, faqat grammatik jihatdan farq qiladi. Masalan, o'zbek tilida "yulduzlar ustida yurmoq" iborasi shodlikda yashamoq ma'nosini ifodalasa, nemis tilida "Auf Wolke sieben schweben" (yetti bulutda suzish) birikmasi ham o'zbek tilidagi muqobili sifatida yuqori kayfiyatni aks ettiradi.

Ko'plab frazeologizmlar xalqning tarixi bilan bog'liq. Masalan, o'zbek tilidagi "otning kallasidek" iborasi hajmning kattaligini bildirish uchun ishlatilsa, nemis tilidagi "so gross wie ein Elefant" (fildek katta) iborasi shunga o'xshash ma'noni anglatadi.

Frazeologik birliklarning ma'nosi, ularning tilda ishlatilgan matniga qarab o'zgaradi. O'zbek va nemis tillarida ba'zi frazeologik birliklarning ma'nosi bir-biriga yaqin bo'lsa ham, ularning qo'llanilishi farq qilishi mumkin.

O'zbek tilida "Bir kunlik qo'shni bo'lmoq" iborasi vaqt o'tishi bilan o'zgarishlar yuz berishi ma'nosini mujassamlashtiradi, nemis tilida esa "Auf der faulen Haut liegen" birikmasi begona ish qilmaslik, uyda qolish ma'nosini anglatadi. Ammo "Bir kunlik qo'shni bo'lmoq" frazeologizmida biror odamning kelajagini oldindan aytish ma'nosi ham mavjud.

Shuningdek, har ikkala tilda ham atoqli shaxslar haqidagi frazeologik birikmalar uchraydi. Masalan, nemis tilida "eynshteyn" (aqli kishi), "nesse" (o'ziga ishonuvchi kishi), o'zbek tilida "shermuhammad" (kuchli odam), "carlson" (chivindiq odam) kabi iboralarni misol qilib keltirish mumkin.

Nemis tilidagi frazeologizmlar o'zbek tilidagi iboralar bilan sinonim ham bo'ladi va ularni qiyosiy tahlil usulida o'rganish mumkin. Masalan, nemis tilidagi "Man empfängt den Mann nach dem Gewand und entlasst ihn nach dem Verstand" frazeologizmi o'zbek tilidagi "Kiyimiga qarab kutib olinadi va aqliga qarab kuzatib qo'yiladi" birikmasi bilan solishtirish mumkin. Biroq, iboralarni qo'llashda, ularning grammatik va leksik jihatdan o'zgarishlar ro'y beradi: "pashshadan fil yasama" iborasi orqali "qo'shib-chatib gapirish" ma'nosi anglashiladi. Bu iboraning nemis tilidagi ekvivalenti esa "Auf einer Mücke einen Elefanten machen".

Tilshunoslikda bir xil ma'noni ifodalovchi turli shakldagi til birliklarining mavjudligi variantlashuv deb nomlanadi. O'zbek va nemis tillarida iboralar variantlashuvga uchrashi mumkin, ya'ni bir ma'noni ifodalash uchun turli usullar qo'llaniladi. Masalan, "Tilingga ehtiyot bo'l" \ "Tilingni tiy" kabi o'zbekcha iboralar hamda "Sei vorsichtig mit deinen Worten" \ "Halte deine Zunge im Zaum" nemischa birikmalarning variantlashuvda iboralar tarkibidagi so'zlar o'zgargan, ammo umumiy ma'no saqlanib qolgan.

Har bir millatning frazeologizmlarida ularning tarixiy, madaniy va ijtimoiy xususiyatlarini aks ettiruvchi realiyalar mavjud. Masalan, o'zbek millatiga xos birliklardan biri mehmondo'stlikdir. O'zbek xalqida mehmon doimiy hurmat bilan kutib olinadi. bundan tashqari, "Osh yeyilguncha, gapirgan yaxshi" — maqoli o'zbek oshxonasiga ishora qilmoqda. Nemis millatida ham o'ziga xos realiyalar bor. Ular aniq rejalashtirish va tartib-intizomga katta ahamiyat berishadi: "Ordnung muss sein" (Tartib bo'lishi shart). Vaqtning juda muhim ekanligi haqida nemis millatiga xos ibora bor: "Zeit ist Geld" (Vaqt - bu pul).

Frazeologizmlarning asosiy tomoni shundaki, bunday birikmalarning bir tildan ikkinchi tilga so'zma-so'z tarjima qilinmasligidadir, chunki iboralarning ma'nosi birikma orasida ishtirok etuvchi so'zlarning ma'nosiga bog'liq emas.

O'zbek va nemis frazeologiyasining farqlaridan biri – bu madaniyat va urf-odatlar o'rtasidagi farqlarni aks ettiradigan frazeologizmning mavjudligidir. O'zbek tilida ko'proq tabiat va oilaviy qadriyatlar, an'anaviy hayotga oid frazeologik birliklar uchraydi. Misol uchun, "Xalqning kuchi qo'lda" iborasi o'zbeklarning mehnatkashligi, qo'l mehnatiga katta ahamiyat berishni ifodalaydi. Nemis tilida esa, asosan, ishbilarmonlikka yo'naltirilgan frazeologik

birliklarni ko'rish mumkin. Masalan, "Den Nagel auf den Kopf treffen" "(muammoni to'g'ri anglash), bu ishning ko'zini bilish hamda ehtiyotkor bo'lishni ifodalaydi.

Nemis va o'zbek tillarining frazeologik fondlari o'ziga xos xususiyatlarga ega, ularni qiyosiy o'rganish orqali quyidagi natijalarga erishish mumkin:

1. Frazeologik birikmalar lug'at boyligining muhim qismi hisoblanadi. Ular orqali tilning ifodaliligi oshadi.
2. Iboralar har bir xalqning mentaliteti, dunyoqarashi va turmush tarzini aks ettiruvchi muhim elementdir.
3. Bir qator frazeologik birliklar qiyoslanayotgan tillarda uchraydi. Bunga sabab ularning qadimda bir tilda bo'lgani yoki bir-biridan o'zlashtirgani bo'lishi mumkin.

Xullas, o'zbek va nemis tillaridagi frazeologik birliklarni qiyosiy tahlil qilish, tillarning o'ziga xos madaniyati, tarixiy tajribasi va ijtimoiy konteksti haqida chuqur tushuncha hosil qiladi. Bu tahlil, o'zbek va nemis tillarini o'rganish jarayonida ikki til orasidagi farqlar va o'xshashliklarni kashf qilish, shuningdek, ularni to'g'ri va samimiy ravishda ishlatishni o'rgatish uchun juda foydalidir.

Foydalanilgan adabiyotlar:

1. Do'sbekov Sherali. Nemis tili so'zlashuv nutqida qiyosiy frazeologik birliklarning ishlatilishi. Toshkent — 2022
2. Imyaminova SH.S. "Nemis tili frazeologiyasi". Toshkent – 2011
3. Abdusaidov A. Frazeologizmlar – matbuot tilida ta'sirchan vosita. Samarqand:2001
4. Rahmatullayev Sh. Frazeologik birliklarning asosiy ma'no turlari. – T.: Akadem nashr. 1985.
5. Mamatov A. O'zbek tili frazeologizmlarining shakllanishi masalalari: Filol. fan. dokt.... diss. avtoref. Toshkent, 2000.
6. 29. Yo'ldoshev B. Hozirgi o'zbek adabiy tilida frazeologik birliklarning funktsional-uslubiy xususiyatlari : Filol. fan. dokt.... diss. avtoref. Toshkent, 1993.
7. Umarxo'jayev M. Olmon tili leksikologiyasi va frazeologiyasi..Andijon 2010
8. Yo'ldoshev B. "Hozirgi o'zbek adabiy tilida frazeologik birliklarning funktsional – uslubiy xususiyatlari : Filol. fan. dokt... diss. avtoref. Toshkent, 1993
9. Shavkat Rahmatullayev."O'zbek tilining frazeologik lug'ati". Toshkent:" O'qituvchi " nashriyoti, 1978.

SYNONYMOUS AND ANTONYMOUS
SERIES OF ADJECTIVES IN THE ENGLISH THESAURUS

Shamiyeva Nargiza

PhD-student of National University of Uzbekistan

Jumayeva Muqaddas

English Teacher of National University of Uzbekistan

Khazratkulov Oybek Ulugbekovich

Teacher of the English Department,
Namangan branch of the Tashkent
International Chemical University

Abstract: synonyms and antonyms are within the study of the significance of adjectives in relation to the same meanings (synonyms) or to completely different meanings (antonyms). The literal interpretation of such relationships remains paramount in analyzing how nuances improve human communication, be it in oral or written contexts. Synonyms and antonyms favor vocabulary enlargement and insight into language intricacy, giving the nugget to educators, writers, and language learners. In simple terms, synonymous adjectives enable speakers and writers to choose words that really carry under a definite context the most precise intention, stressing the importance of subtle effects through changes in their emotional weight and connotation.

Keywords: synonyms, antonyms, adjectives, meanings, analyzing, English thesaurus.

Synonymous adjective series emphasize the significance of minute differences in emotional intensity and connotation by enabling speakers and writers to choose words that express exact meanings within certain circumstances. For example, one may switch “happy” and “joyful”, but the application would differ on the amount of desired feeling by that particular conveyer. On the other hand, antonymous pairs like “hot” and “cold” show a sharp contrast and, thus, adds to the cohesion in discourse, making it richer in description and lively in expression of the ideas at hand. [Jones S., 2012].

Most of the concerns, to put it correctly, have to do with the contextual propriety of such terms and the ambiguity in their meaning. In that case, it becomes easier to confuse meanings through the application of synonymy, which is not context-relevant; while antonyms often provide too simple a choice, failing to capture or misrepresent intricate concepts or emotions. These understandings form quite a significant part of language mastery and its practice, especially in tests like IELTS, where the range of vocabulary used forms part of the parameter for scoring.

Synonyms in context. Synonyms: “big”, “large”, “huge” can be used in different contexts to convey different shades of meaning. Example: “big house”, “large house”, “huge house”. Problem: Using “huge” instead of “big” in a context that implies a normal size can sound unnatural. Adjectives are vital parts of the language that modify nouns, often adding crucial information about their qualities, characteristics, or states. They can show different things like size, color, shape, quantity, etc., which are most useful to describe things in language. To know synonyms, which are words having a more or less similar meaning, is vital for building vocabulary and making communications more exact. This only helps in further enriching vocabulary but also performing discourse cohesion. With the help of synonymous adjectives, the user can make concise expression, avoiding redundancy, and weightiness in terms of message clarity. The systematic arrangement of languages would categorize into certain axes of variations according to how a speaker tends to manipulate communicative choices.

Antonymy and Complexity of Concepts. Antonyms: “hot” and “cold” are simple opposites, but they may not fully convey complex emotions or states. Example: “I feel lukewarm

about this idea” - here “lukewarm” conveys a neutral attitude better than the simple antonym “cold”.

In short, this is a study in synonyms and antonyms in the English thesaurus that would be a stride toward sound awareness of language as well as affecting language learning and processing abilities. With those, a person is now able to communicate more specifically and with depth in meaning, thus showing how the mastery of synonyms and antonyms is sometimes vital in developing communication skills. Antonyms can encompass meanings of any sort. Antonyms, thereby, contribute to the richness and exactitude of the language. Further, the study of antonyms serves the purpose of well-choosing any word that fits the writing best while sharpening the entire language or word awareness. The antonyms can categorize into different classes of antonyms, each characteristically different from the other:

Gradable antonyms are somewhat very far apart. The intensity or quality that applies to the particular pair is involved here. Examples are “hot-cold,” “big-small,” and “fast-slow.” This antonym group amplifies the idea of the characteristic described in terms of variability and relativity.

Complementary antonyms designate pairs of oppositions, thus representing two ends of one dimension. Examples include oppositions like alive-dead, on-off, and male-female. These antonyms have no middle ground: anything is one or the other in a mutually exclusive fashion.

Antonym relations would frequently depend on context to assert how words are construing their opposite meanings in a particular linguistic setting. This relationship helps to invent discourse cohesion while giving a large amount of useful redundancy. Learning antonyms will help enhance one's vocabulary and understanding of the subtleties of the particular language. An understanding of antonyms helps in enhancing explicitness in communication, both orally and written. Antonyms invoke vivid images and precise meanings: The weather today is hot and cold. This helps in creating images that attract much attention.

Role of Thesaurus

Thesaurus enriches language use and writing, honing the fine collection of synonyms and antonyms. Thesaurus derives from the Latin word *thēsauros*, implying “treasure,” or “storehouse,” which mirrors its utility in holding words and meanings. Thesauri mainly serve to help writers find alternative words to substitute repetitive use and embellish their expressions with variety and nuance. For instance, when the writer needs to use a synonym for “happy,” alternatives can result in a full-fledged description from a thesaurus, such as “joyful,” “content,” or “cheerful.” Thesauri are styled differently for distinct audiences and purposes. For example, children's thesauruses typically contain pictures along with simple word lists to aid early readers, while on the other hand, Roget's International Thesaurus would be more appropriate for a professional audience, given its double arrangement that categorizes words according to their meaning and provides serious writers with an extended array of synonyms and nuances. [P.M. Roget,1997].

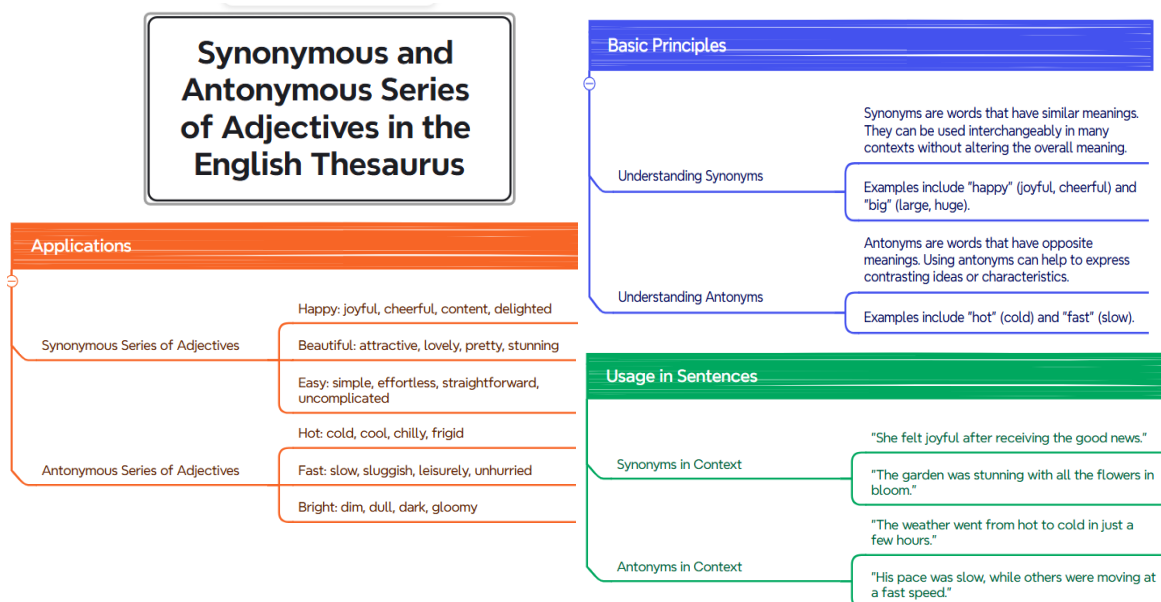
Being printed or otherwise, an online thesaurus can be accessed, providing users at once with vast vocabulary resources. These digital tools are thus most useful in adapting to new language utilizations while helping lexicographers follow the changing meanings.

At the same time, a thesaurus's main design is to have synonymous identification. By using a thesaurus, the writer surveys several synonyms, one of which may serve better to fit his intention. Thus, instead of using the word “big” repeatedly, far better might have been alternatives chosen from a thesaurus like “large,” “enormous,” and “massive.”

Thesaurus also valuably helps locate an antonym, which are those words that are opposite in meaning. This is particularly useful for the writers who wish to bring out contrasting ideas or emotions. For instance, the antonyms for happy, the thesaurus may suggest, are sad and unhappy in a more refined expression of the written word. [Lutz W, 1994].

Thesaurus is also another vital tool for language learners, as they provide understanding about word relationships and vocabulary expansion. Thesauri expose students to slight differences between the meaning and usage that affect successful communication. In print or online,

thesauruses are indispensable if one intends to improve his writing and language identification. [Babajanova L., 2024].



The use of case studies can provide an invigorated perspective in the learning of languages for students. By providing an analytical framework, teachers offer case studies that equip students with enough context and information to explore solutions and their applicability to similar cases. Such approaches encourage systematization within language study; implicating students in issues to be analyzed, goals and contexts, key facts, alternative perceptions, and recommendations. In addition, teachers might expose students to contextualized case studies for the purpose of vocabulary depth; such an approach is very effective when it comes to teaching English as a Foreign Language (EFL).

References

1. Abduraxmonova N. BILINGVAL TEZAVRUS YARATISHDA O 'ZBEK VA INGLIZ TILLARINING VALENTLIK MUNOSABATI. (2024). «CONTEMPORARY TECHNOLOGIES OF COMPUTATIONAL LINGUISTICS», 2(22.04), 454-458.
2. Abdurakhmonova N., Shamieva N. Creating an English-Uzbek Bilingual Thesaurus of Frequently Used Adjectives in Uzbek Corpus //2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE). – IEEE, 2024. – C. 1640-1644.
3. Khamdamova S. B. PECULIAR FEATURES OF WILLIAM BUTLER YEATS' POETRY //Theoretical & Applied Science. – 2020. – №. 4. – C. 348-351.
4. Jones S. Antonyms in English: Construals, constructions and canonicity. – Cambridge University Press, 2012.
5. Dictionary C. E. Thesaurus //URL: <http://dictionary.cambridge.org> (англ.) (дата обращения 05.01.15). – 2006.
6. Thesaurus (lexicography) - Wikiversity
7. Roget P. M. Roget's II: the new thesaurus. – Taylor & Francis, 1997.
8. Lutz W. The Cambridge Thesaurus of American English. – 1994.

SAMARQAND SHEVASI DIALEKTINI TRANSKRIPSIYA QILISH TAMOYILLARI

O‘rinboyeva Ro‘za Toliboyevna

Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti magistranti

E-mail: orinboyevaroza@gmail.com

Asqarov Ilhomjon Jamshid o‘g‘li

Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti magistranti

E-mail: asqarovilhomjon21@gmail.com

Annotatsiya: Ushbu maqolada o‘zbek dialektologiyasi bilan kompyuter lingvistikasi o‘rtasidagi bog‘liqlik ochib berilgan. Jumladan, o‘zbek tili shevalari tasnifi, Samarqand dialektining o‘ziga xos xususiyatlari hamda ularni yangi o‘zbek transkripsiyasida ifodalash jarayoni tahlilga torilgan.

Kalit so‘zlar: dialekt, lahja, transkripsiya, fonetik, leksik, morfologik, audio material.

Abstract. This article reveals the connection between Uzbek dialectology and computer linguistics. In particular, the classification of Uzbek dialects, the specific features of the Samarkand dialect, and the process of their expression in the new Uzbek transcription are analyzed.

Keywords: dialect, dialect, transcription, phonetic, lexical, morphological, audio material.

Аннотация: В статье исследуется связь узбекской диалектологии и компьютерной лингвистики. В частности, проанализированы классификация узбекских диалектов, специфические черты самаркандского диалекта и процесс их выражения в новой узбекской транскрипции.

Ключевые слова: диалект, транскрипция, фонетический, лексический, морфологический, аудиоматериал.

Bugungi kunda kompyuter lingvistikasi bilan dialektologiya fanining o‘zaro uyg‘unligi ko‘zga tashlanadi. O‘zbek dialektologlari tomonidan yaratilgan tadqiqotlar kompyuter lingvistlari tomonidan [Abduraxmonova N., 2021] o‘zbek tili korpusi ostida jamlanmoqda. Bu o‘zbek tili dialektlarini ham kodlash imkoniyatini beradi. Ushbu maqolada Samarqand shevasining o‘ziga xos xususiyatlari aks ettirilib, ularni transkripsiyalash jarayoni batafsil yoritib berilgan.

Manbalarda keltirilishicha, o‘zbek tili shevalarini tadqiq etish Mahmud Koshg‘ariyning “*Devonu lug‘atit turk*” asari orqali boshlanganligi qayd etiladi. Germaniyada “*Yosh grammatikachilar*” tomonidan fan sifatida o‘rganilgan. o‘zbek tilshunosligida bir qator olimlar o‘zbek tili shevalarini tadqiq etishga doir ko‘plab tadqiqotlar yaratgan. Ular o‘zbek tili shevalarini qarluq, qipchoq hamda o‘g‘uz lahjalariga ajratgan holda tasniflaydi. Xususan, professor Samixon Ashirboyevning “*O‘zbek dialektologiyasi*” darsligida ushbu lahjalarning o‘ziga xos xususiyatlari batafsil tahlil etilgan [Ashirboyev S., 2021].

Ayni paytda, o‘zbek tili shevalarini o‘rganish viloyatlar kesimida amalga oshirilmoqda. Xususan, Surxondaryo shevasiga doir A.Turdialiyev, S.Rahimov, M.Hamidov, Buxoro shevasiga doir M.Mirzayev, S.Po‘latova, Qarshi shevasiga doir B.Jo‘rayev, Namangan shevasiga doir A.Aliyev, Xorazm shevasiga doir F.Abdullayevlarning tadqiqot ishlarini misol sifatida keltirish mumkin.

Samarqand shevasini transkripsiyalash uchun, dastavval, uning qaysi lahjaga oidligi hamda uning sifatleri belgilab olinadi. Manbalarda Samarqand viloyatida, asosan, qarluq va qipchoq lahjasi, ayrim hududlarida o‘g‘uz lahjasining qo‘llanilishi qayd etiladi. Quyida bu lahjalarning asosiy xususiyatlari keltirilgan:

1. Qarluq lahjasi.

V.V.Reshetov o'zining "O'zbek shevalarining klassifikatsiyasi" nomli maqolasida Yettisuv, Farg'ona vodiysi hamda Sharqiy Turkiston hududlarida qarluqlar bilan birga chigil, xalaj, yag'mo va uyg'ur qabilalari ham istiqomat qilganligi bois *qarluq-chigil-uyg'ur* lahjasi deb nomlaydi [Reshetov V.V., 1966]. Qarluq lahjasining asosiy farqli jihati shundaki, unda singarmonizmning saqlanishiga ko'ra ikkiga bo'linadi:

– shahar shevalarida singarmonizmga amal qilinmaydi, 6-7 ta unidan foydalaniladi (ä, o, u, a, i, e);

– shimoliy o'zbek va Andijon shevalarida singarmonizm saqlangan holda 9 ta unidan foydalaniladi: a(a), ä, o, ö, u, ü, i, i:(i), e.

Fonetik jihatdan quyidagi xususiyatlar aks etadi:

– qarluq lahjasida o, u, i oraliq unlilari qo'llaniladi;

– e tovushi i kabi, so'z oxiridagi k tovushi y kabi talaffuz qilinadi;

– talaffuzda so'z oxiridagi q va g' tovushlari tushadi;

– so'z oxirida k//g, q//g' almashadi; x//q undoshlari almashinuvi kuzatiladi;

– ä unlisi boshqa shahar shevalaridan farqli ravishda o unlisiga yaqin, ya'ni qisman lablangan holda talaffuz qilinadi: *oldim* => *a(o)llim* kabi.

– talaffuzda quyidagi tovushlar moslashadi:

i-e = ikki ~ ekki, siyrak ~ se:räk;

e-u = devona ~ duvana, devor ~ duval;

e-ö = etik ~ ötü, belanchak ~ bölünčäk;

e-ü = kelin ~ kälän, ekin ~ äkin;

g'-v = tog' ~ tav, bog' ~ bay.

– ko'plik affiksining -lar // lär, -nar // när variantigina qo'llaniladi;

– 6 kelishikli sheva bo'lib, qaratqich kelishigi undoshlardan keyin -in variantlarda qo'shiladi: *otning dumi* – *atin qujrui*, *bizning ishimiz* – *bizin i'şimis*;

– hozirgi zamon davom fe'li -jap affiksi bilan shakllanadi: *boryapman* ~ *barijappan*, *ketyapsan* ~ *ketijapsan*, *oynayapti* ~ *ojnayapti*.

2. Qipchoq lahjasi.

Qipchoq lahjasida eng ko'p uchraydigan fonetik va morfologik xususiyatlar quyidagilar:

– to'liqsiz fe'ldan iborat fe'l shakllarida to'liq singarmonizm mavjud: *bor ekan* ~ *baraqan*, *boribmi ekan* ~ *barippaqan*, *ko'p ekan* ~ *küväkän*;

– so'z boshida j // 3 undoshi mosligi qayd qilinadi: *yo'l* ~ *žol*, *yomon* ~ *žaman*;

– g//y undoshlari j undoshiga o'tadi: *yig'in* ~ *jijin*, *tegdi* ~ *tijdi*, *sigir* ~ *sijir*;

– so'zda unlilar o'rtasida p, k, q undoshlari jaranglashadi: *gäp* ~ *gäbi*, *ekin* ~ *egin*;

– x undoshi q undoshiga o'tadi: *xotin* ~ *qatin*, *tuxum* ~ *tuqum*, *xayol* ~ *qijal*, *xat* ~ *qat*, aksincha, q undoshi x undoshiga ham o'tadi: *vaqt* ~ *vax*, *oq* ~ *ox*;

– so'z oxirida k, q undoshlari tushib qoladi: *sariq* ~ *sari*, *kichik* ~ *kiči*;

– so'z o'rtasida va oxirida z, g', q, k undoshlari v undoshiga o'tadi: *og'iz* ~ *avuz*, *bog'* ~ *bav*, *tog'* ~ *tav*, *bo'yog* ~ *bojav*, *ellik* ~ *eliv*.

3. O'g'uz lahjasi.

O'g'uz lahjasining o'ziga xos sifatlari olimlar tomonidan quyidagicha belgilangan:

– lahja shevalarida kontrast juft unlilar mavjud bo'lib, jami 11 ta unidan foydalaniladi: a, u, ü, i:, i, i:, o, ö, ä, e;

– dastlabki bo'g'inda e unlisi ä unlisiga o'tadi: *kel* ~ *gäl*, *dedi* ~ *dädi*, *berdi* ~ *bärdi*;

– -daj qo'shimchasidagi ä unlisi i unlisiga o'tadi: *shunday* ~ *šundij*, *bunday* ~ *bundij*;

– so'z boshida k tovushi jaranglashadi: *ketti* ~ *gitti*, *kel* ~ *gäl*, *kun* ~ *gün*, *ko'z* ~ *göz*;

– jo'nalish kelishigi -a // ä affikslari bilan ifodalanadi: *otimga* ~ *a:tima*, *ishga* ~ *i:ša*;

– sifatning intensiv shakli uchun *dim* ravishi qo'llaniladi: *dim jaxši*, *dim i:ssi*, *dim jaman*;

– buyruq-istak fe'lining ko'plikdagi I shaxs shakli -(a)li // (a)li affiksi bilan shakllanadi: *kelaylik* ~ *gäläli*, *boraylik* ~ *bäräli*;

– buyruq-istak fe’li II shaxs birligining o’g’uz lahjasi shevalarida o’ziga xos shakli bor: *bersangchi ~ bārsānā, borsangchi ~ barsānā*.

Professor Samixon Ashirboyev o’zining “O’zbek milliy transkripsiyasini yaratish yo’lida” nomli maqolasida lotin alifbosiga asoslangan yangi o’zbek transkripsiyasini joriy etish borasidagi fikrlarini aks ettirgan [Ashirboyev S., 2024: 131-142]. Bugungi kunda dialektologik tadqiqotlar uchun yangi o’zbek transkripsiyasidan foydalaniladi.

Tadqiqot uchun to’plangan audio materiallarni tahlil qilish jarayonida Samarqand viloyatining deyarli barcha hududlarida alohida sheva qo’llanilishiga guvoh bo’ldik. Ularning ba’zilari adabiy tilga yaqin bo’lsa (masalan, Oqdaryo tumani shevasi), ba’zilari butunlay boshqacha (masalan, Qo’shrabot, Urgut tumanlari shevasi). Tadqiqot uchun, asosan, Samarqand viloyati Qo’shrabot tumani o’rozmot, Qoratosh qishloqlari hamda Payariq tumanining Boshqo’rg’on qishlog’i aholisidan yozib olingan audio materiallar to’plangan. Bundan tashqari to’plangan lingvistik baza tarkibida boshqa qishloq shevalari ham mavjud bo’lib, ular juda ozchilikni tashkil etadi.

Inson tilining qaysi shevaga moslashishi, qaysi jihatlarni o’zida shakllantirishi uning yashash sharoiti, muhiti, qolaversa, kasb-koriga ham bog’liq bo’lishi mumkin. Masalan, Samarqand viloyati O’rozmot qishlog’ida tug’ilib-o’sgan, umrining so’nggiga qadar cho’ponlik bilan shug’ullanib kelgan 92 yoshli Musaql otadan yozib olingan audio bilan shu qishloqda istiqomat qiluvchi Azim akadan yozib olingan audioni solishtirganimizda, aynan bitta so’zning ikki xil talaffuz qilinishi, bitta grammatik shaklning ikki xil ko’rinishi, bitta ma’noning ikki xil leksika bilan ifodalanishi kuzatildi. Masalan, 1-audio (Musaql otadan yozib olingan)da “*davr*” so’zi “*dovrim [do:rim]*” tarzida, 2-audioda “*do’rim [dōrim]*” tarzida talaffuz qilingan. Bunda *a* tovushining o’lashishi, ikkinchi misolda *o’* lashishi hamda *r* tovushining tushib qolishi kuzatiladi.

Audiolarni yosh jihatidan tahlil qiladigan bo’lsak, ularning leksikasida farqlarni ko’rish mumkin. Ya’ni katta yoshli (40-50 yoshdan yuqori)lar nutqida asrlar davomida iste’molda bo’lgan shevadagi so’zlar uchraydi. Masalan, “*barikar*” so’zi “*dala*” ma’nosini anglatadi. Tarixiy badiiy asarlarda ham uchratish mumkin. Yoshlar nutqida bo’lsa, zamonaviy leksik qatlam so’zlari faol bo’ladi. Ular, bevosita, zamonaviy texnologiyalarga oid tushunchalarni ifodalash uchun o’zlashma qatlamdan keng foydalanadi. Qolaversa, ularda shahar shevasi elementlari kuzatiladi. Masalan, “*ob-havo ham*” so’zi “*op-hovoyam*” tarzida talaffuz qilinadi. Kasbiy faoliyatidan kelib chiqib tahlil qilinganda, asosan, o’rta ma’lumotli, kolxozchi yoki cho’ponlar nutqida shevaga xos barcha xususiyatlar aks etadi. O’rta ma’lumotli, ta’limdan boshqa soha vakillari nutqida shevaning fonetik hamda morfologik jihatdan muayyan farqlari mavjud bo’lsa, leksik jihatdan adabiy tilga yaqinlashish holati kuzatiladi. Ta’lim sohasi vakillari (o’qituvchilar, maktabgacha ta’lim muassasasi xodimlari kabi) nutqida sheva elementlari deyarli sezilmaydi. Ularda adabiy tilda so’zlash ko’nikmasi nisbatan yuqori bo’lib, ayrim hollardagina o’z shevasidan foydalanishi mumkin (masalan, oilaviy suhbatlarda, do’stlar davrasida).

Samarqand viloyati shevasini transkripsiya qilish jarayonida qarluq hamda qipchoq lahjasi xususiyatlari inobatga olindi. Quyida ularni misollar orqali tahlil qilamiz.

O’rozmot qishlog’i shevasi misolida ko’rib chiqadigan bo’lsak, morfologik jihatdan quyidagi holatlarni kuzatish mumkin:

– tushum kelishigi **-ni** qo’shimchasining **-di/-ti** shaklida qo’llanilishi kuzatiladi:

davrlarni => *dovrlardi [do:vrlardi]*;

bolalarni => *ballardi [bāllārdi]*;

qilishni => *qilishti [qilišti]*;

o’qishni => *o’qishti [ōqišti]*.

– ko’plik shakli yasovchi **-lar** qo’shimchasi **-ar, -sar** shaklida qo’llaniladi:

yaxshimisizlar => *yaxshimissar [jaxšimissär]*;

bizlarning => *bizarding [bizārdin]*.

– qaratqich kelishigining **-ding, -ni, -di, -ti** tarzida qo’llanilishi kuzatiladi:

o’zimning => *o’zimding [özimdin]*;

ollohning => ollodi [ollodi];
 ko 'kni => ko 'kti [kökti];
 sariqni => soriqti [soriqti];
 xabarni => xabardi [xabardi].
 – chiqish kelishi **-dan** qo'shimchasi **-tan, -tin** tarzida qo'llaniladi:
 yoshdan => yashtan [joštän];
 Qodirovning => Qodiroptin [Qodiröptin].
 – II shaxs ko'plik qo'shimchasi **-siz** ning qisqarishi, ya'ni **-s** tarzida qo'llaniladi:
 kelasiz => kelas [kela:s].
 – yubormoq ko'makchi fe'lining **-vormoq, -yarmoq, -jarmoq** tarzida qisqarishi kuzatiladi:
 tashlab yuboring => tashabyaring [tašabjariŋ], tashabjaring [tašabžariŋ].
 – **-ar** sifatdash qo'shimchasi edi, ekan, emish to'liqsiz fe'llari bilan kelganda **-g'ich, -y** shaklida qo'llanilishi kuzatiladi:
 borar edik => barg'ich edik [bäryič edik];
 bo 'lar edi => bo 'laydi [bölaɟdi];
 qiziqar edim => qizig'eydim [qiziɟeydim].
 – I shaxs ko'plik buyruq istak maylining **-eyk** tarzida o'zgaradi: gaplashaylik => gaplasheyk [gaplaše:jk].
 – edi to'liqsiz fe'lining kerak so'zi bilan kelganda qisqarishi kuzatiladi: kerak edi => keragitta [kerägittä].
 – **-mi** so'roq yuklamasi **-ma** tarzida talaffuz qilinadi:
 bormi => borma [bärmä]; charchamayapsizlarmi => charchamapsizarma [čarčamapsizärmä];
 Leksik jihatdan so'zlarning qisqa talaffuz qilinishi yoki so'z tarkibidagi tovushlarning almashishi kuzatiladi. Masalan:
 bilan => binan [binän], minan [minän];
 hozir => haz [hāz], hazir [hāzīr], azir [āzīr];
 lekin => nekin [nekīn];
 obro' => abro'y [äbröj];
 umr => o'mr [ömr];
 ozgina => gartak [gärtäk];
 o'rik => zardoli [zärdoli];
 tok novdasi => taxpilla [täxpilla];
 hammasi => jovlik [žovlik];
 qo'shni => hamsiya [hämsiyä].
 Xulosa qilib aytganda, Samarqand shevasini transkripsiya qilish jarayonida tilning fonetik, morfologik, leksik jihatlari bilan bir qatorda so'zlovchilarning yoshi, jinsi, kasbi hamda ma'lumotiga ham e'tibor berish talab etiladi. Shunga ko'ra audiolarni transkripsiya qilish maqsadga muvofiqdir.

Foydalanilgan adabiyotlar ro'yxati:

1. Ashirboyev S. O'zbek dialektologiyasi (darslik). T.:Nodirabegim, 2021.
2. Решетов В.В. Ўзбек шеваларининг классификацияси // Ўзбек тили ва адабиёти, 1966. 3-14 б.
3. Ashirboyev S. O'zbek milliy transkripsiyasini yaratish yo'lida // Oltin bitiglar - Golden scripts. 2024. 131-142 b. View of O'zbek milliy transkripsiyasini yaratish yo'lida
4. Rahimov S. O'zbek tili Surxondaryo shevalari (fonetika, leksika). T.: Fan, 1985.
5. Mirzayev M. Buxoro oblastidagi o'zbek shevalarini o'rganish uchun anketa. – Buxoro, 1955.
6. Po'latova S. Vobkent tumani shevasining areologik tadqiqi (monografiya). T.: GlobeEdit, 2024.

7. Aliyev A. O'zbek tilining Namangan dialekti bo'yicha material to'plovchilar uchun anketa. – Namangan, 1964.
8. Abdullayev F. Xorazm shevalari. T.: Fan, 1961.
9. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
10. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
11. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzworonet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
12. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus. uz. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.

KO'P KOMPOONENTLI TERMINLARNI PARALLEL KORPUSDA MOSLASHTIRISH MUAMMOLARI INGLIZ VA RUS TILLARI MISOLIDA

Asqarova Madina

Tayanch doktorant

O'zbekiston davlat jahon tillari universiteti

E-mail: mb.asqarova@uzswlu.uz

Annotation: Ushbu maqola ingliz va rus parallel korpusida joylashtirilgan ilmiy matnlardagi ko'p komponentli terminologik birliklarni moslashtirish usuli tahlil qilingan. Parallel korpusdagi matnlarni moslashtirish yondashuvlari, usullari va darajalari tahlil qilingan. Parallel korpusda ko'p komponentli terminlarni moslashtirish uchun ikkita yondashuv tahlil qilingan.

Keywords: Parallel korpus, tema, rema, moslashtirish, sintaktik tuzilma.

Аннотация: В данной статье анализируется метод выравнивания многокомпонентных терминологических единиц в научных текстах, размещенных в англо-русском параллельном корпусе. Проанализированы подходы, методы и уровни выравнивания текстов в параллельных корпусах. Рассмотрены два подхода к выравниванию многокомпонентных терминов в параллельном корпусе.

Ключевые слова: Параллельный корпус, тема, рема, выравнивание, синтаксическая структура.

Abstract: This article analyzes the method of aligning multi-component terminological units in scientific texts placed in the English-Russian parallel corpus. Approaches, methods, and levels of text alignment in parallel corpora are analyzed. Two approaches to aligning multi-component terms in a parallel corpus are examined.

Keywords: Parallel corpus, theme, rheme, alignment, syntactic structure

Parallel korpus korpus lingvistikasining maxsus turi bo'lib har qanday tilda yozilgan asl matn va uning bir yoki bir necha tillarga tarjima qilingan matnlar jamlanmasidir. Parallel korpusning asosiy xususiyatidan bir moslashtirish jarayoni bo'lib, maxsus so'z va terminlarni asl tildagi fragmentlarini hamda maqsad tilidagi matn bo'laklari o'rtasidagi muvofiqlikni aniqlash jarayoni hisoblanadi. Parallel matnlarni moslashtirishda keng qo'llaniladigan metodlar, algoritmlar va dasturiy ta'minotdan foydalanish imkoniyati mavjud bo'lib, ular turli darajadagi ya'ni matn-matn, paragraf -paragraf va gap-gap shaklidagi birliklarini moslashtirish imkonini beradi. [Nosov, 2017; 68]. Matnlarning so'zlar bilan mos kelishi, jumladan bir komponentli atamalar statistik usullar, lingvistik muvofiqlashtirish usullari va so'zlar zanjirini muvofiqlashtirish usullarini o'z ichiga oladi. Parallel matnlarni moslashtirishning statistik usulli bu berilgan ikki parallel matn uchun munosib moslikni topishga qaratilgan. Mazkur usulda moslanayotgan birliklarni hajmi hisobga olinib, moslik darajasi aniqlanadi hamda jumalarning sifati tekshirilganidan so'ng moslashtirish jarayoni amalga oshiriladi. So'zma-so'z moslashtirish usuli-asl matnda qo'llanilgan so'zlarning maqsad tilida ham uchrash ehtimolini ifodalaydi. Har bir moslanayotgan birliklar o'rtasida soddalashtirish amalga oshiriladi, ya'ni asl matn tilidagi har bir so'zga bitta tarjima ekvivalenti beriladi. Biroq bu usul so'z va iboralarni barcha bog'lanishlarini hisobga olmaydi va ko'p komponentli birliklarni moslashtirish jarayoni qo'lda bajarishni talab etadi. Lingvistik usullar grammatik ma'lumotlarga asoslanadi. Masalan, muhim so'z turkumlariga tegishli so'zlarni moslashtirish usuli bo'lib, bunda yordamchi so'z turkumlari inobatga olinmaydi. Ushbu usulni qo'llash uchun parallel korpusning so'z turkumlari bilan belgilangan bo'lishi talab etiladi. Biroq, ushbu usullardan foydalanish barqaror iboralarni yoki realiyalarni moslashtirishda qiyinchiliklarni yuzaga keltiradi. [Asqarova, 2024; 332] Parallel matnlarni semantik moslashtirish transferlar – kognitiv transfer birliklari yordamida amalga oshiriladi, ular asl til va maqsad tilidagi

tuzilmalar o'rtasida funksional va semantik moslikni ta'minlaydi. Ilmiy matnlarni moslashtirishda ko'pincha transkripsiya usuli qo'llaniladi, chunki ko'plab ilmiy atamalar bir xil manbaga ega (yunon, lotin, ingliz yoki fransuz tili). Bu atamalardan tayanch nuqtalar sifatida foydalanish imkonini beradi. Ammo bu usulning asosiy kamchiligi – asl til va maqsad tilidagi so'z tartibining farq qilishidir. [Potemkin, 2012; 138]

Shuni ta'kidlash lozimki, parallel korpusni yaratishda bir komponentli yoki ko'p komponentli atamalarni moslashtirish ba'zi bir muammolarni yuzaga keltiradi. Chunki, ko'p komponentli atamalar 2 dan 12 gacha bo'lgan elementlardan iborat bo'lib, mazkur birliklarning sintaktik, pragmatik, hamda sintagmatik jihatlari tahlil qilinganidan so'ng moslashtirish jarayoni amalga oshiriladi.

Turli formal tuzilmalarga ega bo'lgan terminologik birliklarning parallel matnlarda moslashtirilishi bir qator omillar bilan murakkablashadi:

- Tillar kesimidagi terminologik birliklarning uzunligi va formal tuzilmasining farqliligi;
- Tillar tizimidagi sintaktik tuzilmalarning farqliligi;
- Tarjima transformatsiyalari.

Ko'p komponentli terminlarni moslashtirish usulini yaratishda hisobga olinishi lozim bo'lgan yana bir jihat – taqqoslanayotgan tillarning gap tuzilish ham inobatga olinishi zarur. Misol uchun, ingliz va rus tillaridagi gaplarning strukturasi turli xil bo'lganligi sababli moslashtirish jarayonida ba'zi bir muammolar yuzaga kelishi mumkin.

Gap ikkita asosiy komponentdan tashkil topgan bo'ladi:

- **Tema** – ya'ni, berilgan holatda ma'lum bo'lgan yoki hech bo'lmaganda osongina tushunilishi mumkin bo'lgan qism bo'lib, so'zlovchi o'z fikrini aynan shu qismdan boshlaydi.
- **Rema** – ya'ni, so'zlovchi tomonidan ushbu boshlang'ich nuqta asosida ifodalangan yangi ma'lumotdir.

Gapning mazmuniy bo'linishi turli modellar, chiziqli-dinamik tuzilmalar, so'z tartibi va boshqa vositalar orqali ifodalanadi. Gapdagi tema va rema har qanday gap bo'lagi orqali ifodalanishi mumkin. Rus tilida gapning mazmuniy bo'linishi asosan so'z tartibi va intonatsiya orqali aks ettiriladi. Ingliz tilida esa ifodalash usullari ancha xilma-xildir. Masalan, mavzu-tematik munosabatlar aniq va noaniq artikllar, shuningdek, gapdagi inversiyalar orqali ifodalanishi mumkin. Bundan tashqari, ingliz tilidagi barcha sintaktik tuzilmalarga to'g'ri so'z tartibi xosdir.

Misol uchun :

Qidiruv amalga oshiriladigan hujjatlar guruhini biz kolleksiya deb ataymiz."

Bu holatda mavzu jumlaning birinchi qismida ifodalangan hamda bu jumla tema ya'ni —

"Qidiruv amalga oshiriladigan hujjatlar guruhi", rema esa — **"biz kolleksiya deb ataymiz"**.

Ingliz tiliga tarjimada bu jumla quyidagicha bo'ladi:

"We will refer to the group of documents over which we perform retrieval as a collection",

Terminlarni moslashtirishda e'tiborga olinishi lozim bo'lgan uchinchi omil parallel matnlarda tarjima transformatsiyalarining mavjudligi. Tarjima transformatsiyalari leksik, grammatik va leksik-grammatik transformatsiyalar bo'lib, ular yordamida asl matndan tarjima matniga o'tish amalga oshiriladi. Ingliz va rus tillari juftligida ilmiy-texnik matnlarni tarjima qilishda tarjima transformatsiyalarining keng qo'llanilishi shuni anglatadiki, agar termin qo'shilsa yoki tushirib qoldirilsa, uning ma'nosi boshqa leksik va grammatik vositalar orqali yetkaziladi.

Masalan, quyidagi ruscha gapni ko'rib chiqamiz:

"Интенсивно исследуется новый класс оптических волокон — фотонно-кристаллические волокна, которые расширяют возможности управления дисперсионными характеристиками и нелинейными свойствами волокон."

Uning ingliz tilidagi tarjima ekvivalenti:

“A lot of research is being carried out regarding a new class of optical fibers, photonic crystal fibers, which extend the ability to manage the dispersion characteristics and nonlinear properties of fibers.”

Bu yerda исследуется fe’li research termini bilan ifodalangan, shuningdek, управление termini inglizcha manage fe’li orqali tarjima qilingan. Bu misol leksik-grammatik transformatsiyalar orqali so’z shakllari va ularning grammatik kategoriyalaridagi o’zgarishlar orqali ma’no yetkazilishining namunasidir.

Shunday qilib, matn moslashtirish jarayonida asl va tarjima matnlarida terminologik birliklarning soni bir xil bo’lmazligi mumkin. Bu tafovut, asosan, tarjima transformatsiyalarining qo’llanilishi, terminlarning qo’shilishi yoki tushirib qoldirilishi, shuningdek, leksik va grammatik moslashuvlar tufayli yuzaga keladi. [Krylova, 2012; 7]. Ilmiy va texnik matnlarni parallel korpusda moslashtirish matn-matn, paragraf-paragraf va gap-gap darajalarida amalga oshiriladi. Parallel matnlarni moslashtirishning eng keng tarqalgan usullari quyidagilardan iborat: statistik usullar, lingvistik moslashtirish usullari, so’zlar satrlarini moslashtirish usullari hamda neyron tarmoqlardan foydalanishga asoslangan usullar.

Ko’p komponentli terminlarni parallel matnlarda moslashtirish bir qator murakkabliklarni keltirib chiqaradi, jumladan:

- terminologik birliklarning formal uzunliklarining farqlanishi,
- ingliz va rus tillaridagi sintaktik tuzilmalarning o’zaro farqliligi,
- tarjima transformatsiyalarining keng qo’llanilishi.

Ko’p komponentli terminlarni moslashtirish uchun ikki xil ssenariy ko’rib chiqiladi:

1. **So’z bo’yicha moslashtirish**, so’ngra ingliz va rus tillaridagi terminlarning sintaktik andozalariga asoslangan holda terminlarni ajratib olish.
2. **"Terminlar to’plami" (bag of terms) shakllantirish**, bu usulda ingliz va rus tillaridagi sintaktik andozalarga asoslangan holda terminlar ro’yxati tuziladi va keyin ushbu ro’yxatlar moslashtiriladi. [Zakharov, 2015; 22]

So’zlarni moslashtirish jarayonida so’z tartibi qat’iy bo’lgan tillar juftligida yuqori samaradorlik kuzatiladi.

Foydalanilgan adabiyotlar:

1. Nosov, A.V., Linguistic annotation of the parallel corpora, Industriya Perevoda, 2017, vol. 1, pp. 68–72.
2. Asqarova, M., “Linguistic problems of text annotation in corpus linguistics”. 2024. N 4, Science problem.uz ISSN:2181-1342.
3. Potemkin, S.B., Problems of developing a parallel corpus of translations of Russian classics, Armiya O-vo., 2012, no. 2, pp. 138–146
4. Krylova, O.A., Structural schemes and actual division of a sentence, Vestn. Ross. Univ. Druzhby Narodov, Ser.: Teoriya Yazyka. Semiotika. Semantika, 2012, no. 2, pp. 6–14
5. Zakharov, V.P., Corpora of the Russian language, Tr. Inst. Russ. Yazyka V.V. Vinogradova, 2015, vol. 6, pp. 20–65.
6. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
7. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).

NER: TEGGLASH MUOMMOLARI

Elov Botir Boltayevich

Alisher Navoiy nomidagi

Toshkent davlat o'zbek tili va adabiyoti universiteti.

Texnika fanlari falsafa doktori(PhD), dotsent

E-mail: elov@navoiy-uni.uz

ORCID: 0000-0001-5032-6648

Samatboyeva Madina To'liqinjon qizi

Alisher Navoiy nomidagi

Toshkent davlat o'zbek tili va adabiyoti universiteti.

Kompyuter lingvistikasi va raqamli

texnologiyalar kafedrası tayanch doktoranti.

E-mail: samatboyevamadina@navoiy-uni.uz

ORCID: 0000-0002-5146-4484

Annotatsiya: Tabiiy tilni qayta ishlash (NLP) sohasida "Named Entity Recognition" (NER) — ya'ni nomlangan obyektlarni tanish — tilshunoslik va kompyuter ilm-fanining muhim va murakkab yo'nalishlaridan biridir. NER, odatda, matn ichida shaxsiy nomlar, joy nomlari, tashkilot nomlari, vaqt va boshqa maxsus terminlarni aniqlash va teglar qo'yish jarayonini anglatadi. O'zbek tili kabi morfologik jihatdan boy va sintaktik jihatdan erkin tilda NER tizimlarini ishlab chiqish va ularni samarali tarzda qo'llash o'ziga xos qiyinchiliklarga ega. Ushbu maqolada O'zbek tilida NER teglashning muammolari, imkoniyatlari va yechimlari haqida so'z yuritiladi.

Kalit so'zlar: NER, NLP, atoqli ot, nomlangan obyekt, identifikatsiyalash, klassifikatsiyalash, teg.

Аннотация: В области обработки естественного языка (NLP) распознавание именованных существей (Named Entity Recognition, NER) — это один из важных и сложных направлений лингвистики и компьютерных наук. NER обычно означает процесс выявления и разметки в тексте имен собственных, географических названий, названий организаций, временных выражений и других специальных терминов. Разработка и эффективное применение NER-систем для узбекского языка, обладающего богатой морфологией и свободным синтаксисом, представляет собой особые трудности. В данной статье рассматриваются проблемы, возможности и решения в области разметки NER для узбекского языка.

Ключевые слова: NER, NLP, собственное имя, именованная сущность, идентификация, классификация, тег.

Abstract: In the field of Natural Language Processing (NLP), Named Entity Recognition (NER) is one of the important and complex areas of linguistics and computer science. NER generally refers to the process of identifying and tagging proper names, location names, organization names, time expressions, and other specific terms in a text. Developing and effectively applying NER systems for a morphologically rich and syntactically free language like Uzbek presents unique challenges. This article discusses the problems, opportunities, and solutions related to NER tagging in the Uzbek language.

Keywords: NER, NLP, proper noun, named entity, identification, classification, tag.

NER, tabiiy tilni qayta ishlashning asosiy vazifalaridan biri bo'lib, u matndagi muayyan nomlangan obyektlarni (shaxslar, joylar, tashkilotlar va boshqalar) aniqlashni maqsad qiladi. Named Entity Recognition (NER) — bu matndagi ma'lumotlarni aniqlash, belgilash va

kategoriyalash jarayonidir. NER tabiiy tilni qayta ishlash (NLP) ning bir shakli bo'lib, bu mashinalarga tabiiy tillarni analiz qilish va qayta ishlash imkonini beradi.

NER (Nominal Entity Recognition) matndan ma'lumotlarni aniqlaydi va foydalanuvchiga soddalashtirilgan formatda taqdim etadi. Bu tibbiyot, marketing, jurnalistika va inson resurslari kabi sohalarda ko'plab qo'llaniladi.

NERlarni avtomatik aniqlash jarayonida til grammatik qoidalariga asoslanadi. Modellar, eng avvalo, NER obyektlarining, asosan, ot so'z turkumiga xosligini ko'rib chiqadi. Undan so'ngra ot so'z turkumiga xos leksik hamda grammatik qoidalar o'rganiladi. Masalan, ot so'z turkumi doirasida NERlar atoqli otlarni ifodalaydi. Atoqli otlar matn tarkibida hamisha bosh harf(capital letter) bilan yoziladi. Bu NERlarni avtomatik aniqlovchi model uchun asosiy qoidalarni ishlab chiqishda yordam beradi.

Tadqiqotchilarning fikricha, har qanday tildagi atoqli otlar soni o'sha tildagi oddiy so'zlarning miqdoriga nisbatan bir necha bor ortiq. Atoqli otlar odatda bosh harf bilan yoziladi ammo arab grafikasiga asoslangan tarixiy yozuvlarda hatto lotin grafikasi asosidagi o'zbek yozuvining dastlabki davrlarida atoqli otlar boshqa oddiy so'zlar qatori kichik harf bilan yozilgan [Xusainova. 2017; 13].

NER tizimlari asosan ikki bosqichda ishlaydi:

1. **Identifikatsiya (Identifying)** – Matnda nomlangan obyekt(NER)larni topish.
2. **Klassifikatsiya (Classifying)** – Topilgan obyekt(NER)larni to'g'ri kategoriyalarga ajratish, masalan, shaxs, joy, tashkilot va h.k.

Nomlangan obyektlar (NER)ni tanib olish va uning turlarini avtomatik identifikatsiya qilish vazifasi axborot olishning eng intensiv o'rganilgan usullaridan biridir. Til tahlilining taqdimoti odatda lug'atdagi so'zlarni qidirish va ularni ot, fe'l, sifat va boshqalar sifatida aniqlashdan boshlanadi. Lekin ko'pchilik matnlar juda ko'p ismlarni o'z ichiga oladi. Agar tizim ularni lug'atdan topa olmasa, ularni aniqlay olmaydi. Bu o'rinda matnni lingvistik tahlil qilish qiyin. Shunday qilib, NER ko'plab tabiiy tillarni qayta ishlash (NLP) vazifalarida (masalan, ma'lumot olish (IR) yoki mashina tarjimasi (MT)) muhim ahamiyatga ega. [Simon. 2013; 12].

O'zbek tilida NER teglashning muammolari

Matnni teglash/annotatsiyalashda matn ma'lumotlarini turli xil tabiiy tillarni qayta ishlash vazifalari uchun tushunarli va foydalanishga imkon beradigan yorliq(metka)lar yoki toifalarni belgilash jarayonidir [Yodgorov. 2024; 149]. Teg – Kompyuter yordamida matn tahlilini amalga oshirish jarayonini tezlashtirish va osonlashtirishga xizmat qiluvchi shartli belgi yoki maxsus kod [Elov, Hamroyeva, Abdullayeva, Uzoqova. 2022; 52].

Matn tarkibidan NER obyektlarini avtomatik aniqlashda bir nechta muammolar mavjud. Bu muammolar matn tarkibidagi NER obyektlarini avtomatik tanib olish, ularga mos modellar yaratish, teglash, obyekt turini hamda uning matn tarkibidagi chegarasini belgilashda yuzaga keladi [Samatboyeva M. 2022; 112].

Morfologik murakkabliklar

O'zbek tilidagi morfologik xususiyatlar, masalan, otlarning ko'pligi, otlarning shakllari va qo'shimchalar, nomlangan obyektlarni aniqlashda qiyinchilik tug'dirishi mumkin. Morfologik tahlil jarayonida bir qancha murakkabliklar yuzaga keladi:

So'zlar turli qo'shimchalar orqali ko'plab shakllarga kirishi mumkin. Masalan, "Samarqand" so'zi O'zbekistondagi tarixiy shaharni bildirsa, "samarqandlik" so'zi esa odamni anglatadi. Bunda so'zga qo'shimcha qo'shilgach, uning matn tarkibidagi ma'nosi ham o'zgaradi. Avvaliga NER bo'lgan (Samarqand) so'z, endi unday hisoblanmaydi (samarqandlik).

Morfologik tahlilda omonimiya. Masalan "Uzun" (ot) - Uzun tumani (Surxondaryo viloyati) va "uzun" (sifat) – "kalta" so'zining antonimi. Bunday shakldosh holatlarda so'zning turkumi juda muhim chunki bu nafaqat NERni aniqlash, balki unga to'g'ri toifa berishni ham anglatadi.

Shunday qilib, nomlangan obyektlarni to'g'ri ajratish uchun so'zning shakli va mazmuni to'liq tahlil qilinishi zarur.

Matndagi ma'no (Kontekstual semantika).

O'zbek tilidagi ba'zi so'zlar bir-biriga o'xshash fonetik tuzilishga ega bo'lishi mumkin. Bu tilda bir nechta ma'nolarga ega so'zlar mavjudligini ko'rsatadi. Masalan, "O'zbekiston" so'zi bir tomondan davlat nomini anglatishi mumkin, ikkinchi tomondan esa "O'zbekiston" nomli mehmonxonani anglatishi mumkin. Bu so'zlarning matn tarkibidagi ma'nolari turlicha. Tilimizda esa bunday birliklar juda ko'p: (Qarang: 1- jadval).

N ^o	Joy nomlari	Shakldoshlik holati
1	Zarafshon	1. Vodiy nomi (Samarqand va Qashqadaryo hududlari) 2. Daryo nomi 3. Shahar nomi (Navoiy viloyati) 4. Ko'cha nomi (Navoiy viloyati) 5. Tog' tizmasi nomi (Pomir-Oloy tizmalari)
2	Sirdaryo	1. Viloyat nomi 2. Daryo nomi 3. Telekanal nomi
3	Andijon	1. Viloyat nomi 2. Shahar nomi 3. Suv ombori 4. Mehmonxona nomi 5. Telekanal nomi 6. Muzqaymoq nomi 7. Polka nomi
4	Samarqand	1. Viloyat nomi 2. Shahar nomi 3. Do'kon nomi 4. Mehmonxona nomi 5. Kafe nomi 6. Restoran nomi 7. "Samarqand" hududgaz ta'minoti MChJ 8. "Samarqand" asari (prof. Z. M. Akramov) 9. Gazeta nomi 10. Telekanal nomi 11. "Samarqand" non

1-jadval. NER("Location" turi – joy nomi) obyektlarining omonimlik hosil qilishi

Resurslarning yetishmasligi. O'zbek tilida NER tizimlarini yaratishda ko'p hollarda tegishli korpuslar, annotatsiyalar va tahlilchilar kerak bo'ladi. Shuningdek, boshlang'ich baza (database) uchun lug'atlar tizimi muhimdir. O'zbek tili uchun NERlarni aniqlashda atqoli otlar asosida tuzilgan lug'atlar bazasi asosiy manba bo'lib xizmat qiladi. Inson ismlari lug'ati, joy nomlari lug'ati shular jumlasidandir.

Nomlangan obyektlarning shakllari. O'zbek tilida nomlangan obyektlar turli shakllarda bo'lishi mumkin. O'zbek nomlari ko'p hollarda qo'shimchalar orqali farqlanadi, masalan: shaxsiy nomlar (Alisher Omonov), joy nomlari (Toshkent), tashkilot nomlari (Navoiy metrostansiyasi), va h.k. Shu bilan birga, ular ko'pincha birgalikda ishlatiladi, masalan, "Alisher Navoiy universiteti". Bunday holatlar tizimni murakkablashtiradi. Bunda tokenlar orqasidagi bog'lanish va farqlanish muhimdir.

Xulosa. O'zbek tilida NER teglash tizimlarini ishlab chiqish va rivojlantirish katta imkoniyatlarga ega, lekin bir qancha muammolarni ham o'z ichiga oladi. Morfologik va sintaktik xususiyatlar, kontekstni to'g'ri tahlil qilish, resurslarning yetishmasligi kabi qiyinchiliklarga qaramay, ilg'or texnologiyalar va innovatsion yondashuvlar yordamida bu masalalarni hal qilish

mumkin. NER tizimlarini takomillashtirish va yangi resurslar yaratish O'zbek tilini kompyuter yordamida qayta ishlash sohasida katta yutuqlarga erishishga imkon beradi.

Foydalanilgan adabiyotlar:

1. Xusainova L. Atoqli otlarning leksik-grammatik guruhlari. Magistrlik dissertatsiyasi. Urganch. 2017. 13b
2. Simon, Eszter. 2013. Approaches to Hungarian Named Entity Recognition. PhD. Thesis. P-12. http://www.cogsci.bme.hu/~ktkuser/PHD_iskola/dissertations/20131011_Simon_Eszter/ertekezes.pdf
3. Samatboyeva M. 2022. Matnlarda NER obyektlarini aniqlashdagi muammolar. O'zbekiston: til va madaniyat. Amaliy filologiya. 2 (5): 110-118.]
4. Yodgorov U. Korpus matnlarini teglash muammolari. Kompyuter lingvistikasi: muammolar, yechim, istiqbollar xalqaro ilmiy-amaliy konferensiya to'plami. Vol. 1 №. 01 (2024). 147-154.
5. Elov B, Hamroyeva Sh, Abdullayeva O, Uzoqova M. O'zbek tilida pos tegging masalasi: muammo va takliflar. O'zbekiston: til va madaniyat. Amaliy filologiya. 2 (5): 51-68.

SHAXS-SON QO'SHIMCHALARINI AVTOMATIK POS-TEGLASH

Erkinjonova Go'zal Baxtiyor qizi

Alisher Navoiy nomidagi

Toshkent davlat o'zbek tili va adabiyoti universiteti

Kompyuter lingvistikasi fakulteti 1-kurs magistranti

E-mail: uzbekistan-24@mail.ru

Annotatsiya: Shaxs-son qo'shimchalarini avtomatik pos teglash tizimi tilshunoslikning turli sohalarida, korpus lingvistikasida, matn tahlili, tabiiy tilni qayta ishlashda muhim ahamiyatga ega. Maqolada shaxs-son qo'shimchalarini morfologik xususiyatlari, ularni aniqlashda lingvistik yondashuvlar, avtomatik pos teglashni amalga oshirish uchun zarur bo'lgan algoritmlar va modellar yoritilgan.

Kalit so'zlar: shaxs-son, morfoanalizator, affiks, POS-teglash, morfologik tahlil.

Abstract: The person-number affixes in automatic POS tagging systems are of significant importance in various fields of linguistics, corpus linguistics, text analysis, and natural language processing. This article discusses the morphological characteristics of person-number affixes, linguistic approaches to identifying them, and the algorithms and models necessary for implementing automatic POS tagging.

Keywords: person-number, morfoanalyzer, affix, POS-tagging, morphological analysis.

Аннотация: Аффиксы лица и числа в системах автоматической разметки частей речи имеют большое значение в различных областях лингвистики, корпусной лингвистики, анализа текстов и обработки естественного языка. В статье рассматриваются морфологические характеристики аффиксов лица и числа, лингвистические подходы к их определению, а также алгоритмы и модели, необходимые для реализации автоматической разметки частей речи.

Ключевые слова: лицо-число, морфоанализатор, аффикс, POS-теггинг, морфологический анализ.

Kirish

Tabiiy tilda ma'lumotlarni qayta ishlash hozirgi kunda kompyuter lingvistikasi uchun dolzarb va kerakli soha sifatida rivojlanmoqda. O'zbek tili, boshqa turk tili singari, morfologik jihatdan o'zgacha xususiyatga ega bo'lib, uning asosiy xususiyatlaridan biri so'z yasovchi va infleksiyalashuvchi affikslarning kengayishidir. Ayniqsa fe'l so'z turkumining grammatik kategoriyalari so'zning grammatik ma'nosini aniqlashda muhim ro'l o'ynaydi.

Turk tilshunosligida shaxs-son kategoriyasini ko'plab olimlar tadqiq etgan. [Ozer, N. A., 2012] turk tilida shaxs va son qo'shimchalarining morfologik strukturasi, ularning ishlatilishini ko'rib chiqqan.[Yilmaz, A. B.,2005.] fe'l qo'shimchalarining shakllarini o'rganib, shaxs-son qo'shimchalarining turli sintaktik va semantik vazifalarini ko'rsatgan.

O'zbek tilshunosligida ham korpus lingvistikasi bo'yicha ko'plab amaliy ishlar olib borilmoqda. Sh.Hamroyevaning "O'zbek tili morfologik analizatorining lingvistik ta'minoti" mavzusidagi dissertatsiyasida o'zbek tili morfoanalizatori uchun grammatik kategoriyalarni o'zbek tili qonuniyatlari asosida qat'iy belgilash lozimligini ta'kidlab, shaxs-son qo'shimchalar uchun grammatik qoidalar keltirib o'tgan. [Hamrayeva, 2021:150]. "O'zbek tilining morfologiyasi" da fe'lning morfologik tizimi va fe'l qo'shimchalarining xususiyatlari keng talqin qilingan. [Karimov, Sh , 2005.].

Asosiy qism

Avtomatik pos-teglash (part-of-speech tagging) tizimi soʻzning turini va uning grammatik xususiyatlarini aniqlashda qoʻllaniladi. Bu tizimning vazifasi shaxs-son qoʻshimchalarini aniqlash va ularga tegishli grammatik xususiyatlarni belgilashdir.

Morfologik tahlil: Shaxs-son qoʻshimchalarini aniqlashda morfologik analizatorlar yordamida soʻzlarning asosi va qoʻshimchalarini ajratish amalga oshiriladi.

Oʻzbek tilida shaxs-son qoʻshimchalari 3 turga boʻlinadi:

1-tur

Birlik			Koʻplik		
I	-m	Prs1s	I	-k	Prs1p
II	-ng	Prs2s	II	-ingz	Prs2p
III	0	Prs3s	III	0	Prs3p

2-tur

Birlik			Koʻplik		
I	-man	Prs1s	I	-miz	Prs1p
II	-san	Prs2s	II	-siz	Prs2p
III	-di yoki 0	Prs3s	III	-di (lar) yoki 0	Prs3p

3-tur

Birlik			Koʻplik		
I	-a(y), -a(yin)	Prs1s	I	-a(ylik)	Prs1p
II	-gin	Prs2s	II	-(i)ng	Prs2p
III	-sin	Prs3s	III	-sin(lar)	Prs3p

Morfoanaliz uchun Sh.Hamrayeva oʻzining Oʻzbek tili morfologik analizatorining lingvistik taʼminoti” mavzusidagi dissertatsiyasida quyidagi koʻrinishda bergan.

1. V + f1v + r1f. Bunda: f1v = {-gan /-(i)b / -yap / -a / -y / -moqchi}; r1f = {-man / -san / -0 / -miz / -siz / -(di)lar};
2. V + f2v + r2f. Bunda: f2v = {-di / -sa}; r2f = {-m / -ng / -0 / -k / -ngiz / -(di)lar};
3. V + r3f. Bunda: r3f = {-ay / -0 / -sin / -aylik / -gin / -(i)ngiz / -(sin)lar / +ishsin}.

Oʻzbek tili agglunativ tillarga kirishi sababli, koʻpincha, lemma stemga toʻgʻri keladi. Feʼning shaxs –son qoʻshimchalarini aniqlashda ham flektiv tillardagi kabi murakkab boʻlmaydi. Morfologik tahlil oʻgdan chapga va chapdan oʻnga qoidasi orqali ishlaydi.

Morfologik tahlilning uchta bosqichi mavjud: 1. Stemlash- soʻzning asosini aniqlash. 2. Lemmalash – soʻzni lemmaga keltirish. 3. Grammemalash- soʻzning grammatik shakllarini aniqlash jarayoni.

Shaxs-son qoʻshimchalarini aniqlashda bir necha usullardan foydalanish mumkin. *Affikslar lugʻati yordamida tahlil qilish* usuli orqali soʻzdan affikslarni ajratib olib, lugʻatdan qidirish; *Lugʻat yordamida tahlil qilish usuli* orqali soʻz formasini topish va unga mos keladigan grammatik maʼno qaytariladi. Shaxs-son qoʻshimchalarini lugʻatlar orqali keltirishda soʻzning lemmasi va qoʻshimchalar qanday oʻzgarishini tahlil qilish mumkin. Bu tahlil koʻproq infleksiyali til uchun moʻljallangan, agglunativ tillar uchun emas.

Shaxs-son qoʻshimchalarini avtomatik aniqlashda qoidalarga asoslangan tizimlar va statistik modellar ham samaralidir. *Qoidalaarga asoslangan tizimda* shaxs-son qoʻshimchalarini uchun aniq qoidalar ishlab chiqiladi. Har bir soʻzni qoidalarga asoslanib aniqlaydi. *Statistik modellar esa n-gram* modellaridan foydalanib, matndagi soʻzlarning kelishiga qarab shaxs-son qoʻshimchalarini aniqlaydi. Shaxs-son qoʻshimchalarini aniqlashda biz ushbu yoʻnalishdan foydalandik;

Tokenizatsiya- matn alohida soʻzlarga ajratiladi;

Morfologik segmentatsiya- feʼl asosi qoʻshimchalardan ajratiladi;

Qo'shimchalarni moslashtirish- so'z oxiri oldindan belgilangan qo'shimchalar ro'yxati bilan taqqoslanadi;

Teglash- tegi aniqlanadi.

Qo'shimchalar to'plami – dastur o'zbek tilidagi shaxs-son qo'shimchalarini lug'at shaklida saqlaydi. Matnni so'zlarga ajratish – foydalanuvchi kiritgan matn *split()* funksiyasi yordamida so'zlarga ajratiladi. Shaxs-son qo'shimchalarini aniqlash – har bir so'z o'zining oxirgi qismidagi qo'shimchalarga qarab tekshiriladi. Agar so'z ma'lum bir shaxs-son qo'shimchasi bilan tugasa, u natijaga qo'shiladi. Natijani chiqarish – topilgan shaxs-son shakllari foydalanuvchiga ko'rsatiladi. Misol uchun, agar foydalanuvchi quyidagi matnni kiritrsa:

Kiritish:

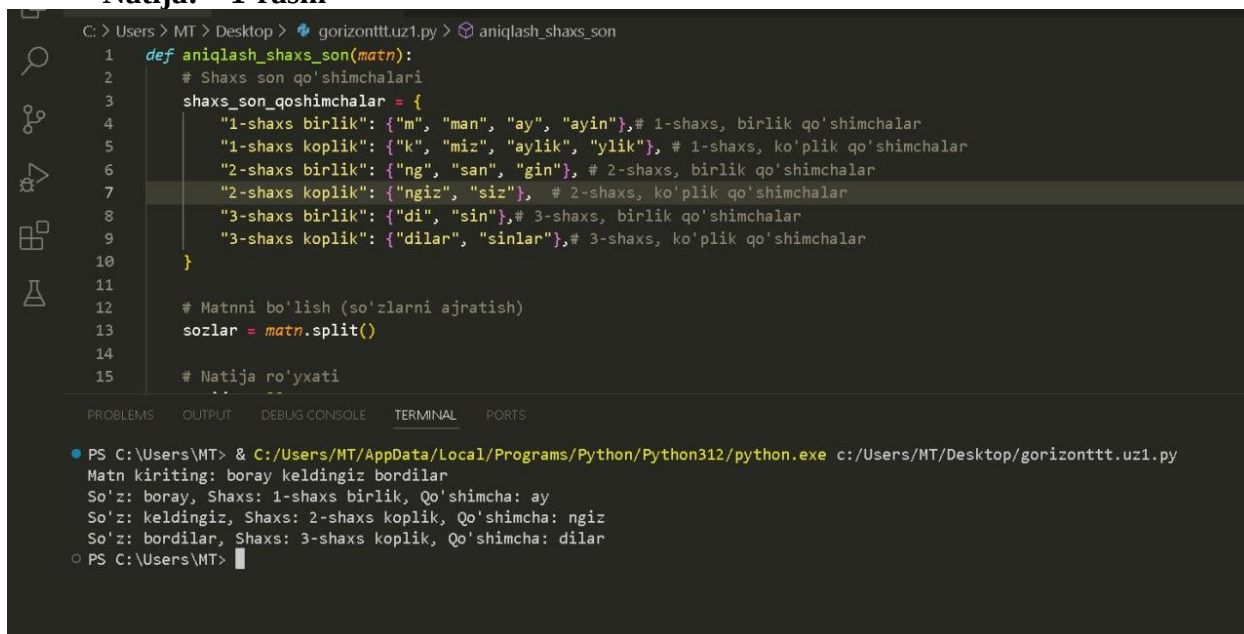
Men kitob o'qiyapman, sen ishlayapsan, ular yozishyapti.

Natija:

So'z: o'qiyapman, **shaxs:** 1-shaxs birlik, **qo'shimcha:** man

So'z: ishlayapsan, **shaxs:** 2-shaxs birlik, **qo'shimcha:** san

Natija: 1-rasm



```

C: > Users > MT > Desktop > gorizonttt.uz1.py > aniqlash_shaxs_son
1 def aniqlash_shaxs_son(matn):
2     # Shaxs son qo'shimchalari
3     shaxs_son_qoshimchalar = {
4         "1-shaxs birlik": {"m", "man", "ay", "ayin"}, # 1-shaxs, birlik qo'shimchalar
5         "1-shaxs koplik": {"k", "miz", "aylik", "ylik"}, # 1-shaxs, ko'plik qo'shimchalar
6         "2-shaxs birlik": {"ng", "san", "gin"}, # 2-shaxs, birlik qo'shimchalar
7         "2-shaxs koplik": {"ngiz", "siz"}, # 2-shaxs, ko'plik qo'shimchalar
8         "3-shaxs birlik": {"di", "sin"}, # 3-shaxs, birlik qo'shimchalar
9         "3-shaxs koplik": {"dilar", "sinlar"}, # 3-shaxs, ko'plik qo'shimchalar
10    }
11
12    # Matnni bo'lish (so'zlarni ajratish)
13    sozlar = matn.split()
14
15    # Natija ro'yxati
16    ...
17
18    PS C:\Users\MT> & C:\Users\MT\AppData\Local\Programs\Python\Python312\python.exe c:/Users/MT/Desktop/gorizonttt.uz1.py
Matn kiriting: boray keldingiz bordilar
So'z: boray, Shaxs: 1-shaxs birlik, Qo'shimcha: ay
So'z: keldingiz, Shaxs: 2-shaxs koplik, Qo'shimcha: ngiz
So'z: bordilar, Shaxs: 3-shaxs koplik, Qo'shimcha: dilar
    
```

Shaxs-son qo'shimchalarini avtomatik aniqlashda ba'zi qiyinchiliklar mavjud. O'zbek tili zamon shakllari va shaxs-son shakllarida sinonimiya kuzatiladi. –*di* qo'shimchasi o'tgan zamon va III shaxs bo'lib kelishi mumkin. (bordi- o'tgan zamon qo'shimchasi, III shaxs birlik shaxs-son 0 shaklda kelgan)

–*gin* qo'shimchasi buyruq-istak mayli va II shaxs shaxs-son vazifasida ham kelishi mumkin. Shu kabi sinonimlik qo'shimcha aniqlashda murakkablikni keltirib chiqarmoqda,

Xulosa

Shaxs-son qo'shimchalarini avtomatik pos-teglash o'zbek tilini qayta ishlash tizimlarida muhim o'rin tutadi. Maqolada tasvirlangan tizim, morfologik tahlil va sintaktik yondoshuvlarni birlashtirib, so'zlar va ularning grammatik shakllarini aniqlashda samarali ishlash imkonini beradi. Tizimning samaradorligi real matnlar bo'yicha testlardan o'tkazilgan va u dastlabki analiz jarayonida yuqori aniqlikni ta'minladi. Kelajakda, ushbu tizimni yanada takomillashtirish va uning turli til modellari bilan kengaytirilishi mumkin.

Foydalanilgan adabiyotlar ro'yhati:

1. Хамроева III. Ўзбек тили морфологик анализаторининг лингвистик таъминоти: Филол. фан. док. ...дис. – Тошкент, 2021. – Б. 150.

2. Hamroyeva Sh., Xoliyoyorov O‘., Abdualimova G. O‘zbek tilining fe‘l shakllari lug‘ati. (“kiymoq” fe‘li misolida) I jild. – Termiz, 2021. – B. 19.
3. Elov B., Hamroyeva Sh., Abdullayeva O., Uzakova M. O‘zbek tilida POS tegging masalasi: muammo va takliflar / Uzbekistan: Language and Culture. Applied philology. 2022/2(5). – 51-68-b.
4. Ғуломов А. Феъл. – Тошкент: Фан, 1954. – Б. 19–38
5. Karimov, Sh. "O'zbek tilining morfologiyasi", 2005.
6. Shodmonov, Sh. "O'zbek tili fe'lining morfologiyasi", 2014
7. https://www.researchgate.net/publication/361114733_Morpholex_Turkish_A_Morphological_Lexicon_for_Turkish
8. <https://www.barandergisi.net/turk-dilinin-morfolojisi>
9. <https://ucrel.lancs.ac.uk/papers/ClawsWordTaggingSystemRG87.pdf>
10. <https://uznatcorpara>

“ABDURAUF FITRAT KORPUSI” MATNLARINI LEKSIK QATLAMGA KO‘RA TEGGLASH

Hamroyeva Shahlo Mirjonovna

filologiya fanlari doktori, professor
Toshkent davlat o‘zbek tili va adabiyoti universiteti

Alimbekov Mavjuda Xalimjon qizi

Toshkent davlat o‘zbek tili va adabiyoti universiteti
tayanch doktoranti

Annotatsiya: Ushbu maqolada “Abdurauf Fitrat korpusi” matnlarining lemmalarini yarimavtomat usulda aniqlash va leksik qatlamga ko‘ra teglash jarayoni yoritilgan.

Kalit so‘zlar: Abdurauf Fitrat korpusi, lemma, leksik qatlam, o‘z qatlam, o‘zlashma qatlam, teglash.

Abstract: This article describes the process of semi-automatic identification of lemmas of the texts of the “Abdurauf Fitrat corpus” and tagging them according to lexical layer.

Keywords: Abdurauf Fitrat corpus, lemma, lexical layer, own layer, borrowed layer, tagging.

Аннотация: В данной статье освещен процесс полуавтоматического обнаружения лемм текстов «Корпуса Абдурауфа Фитрата» и их разметки по лексическому слою.

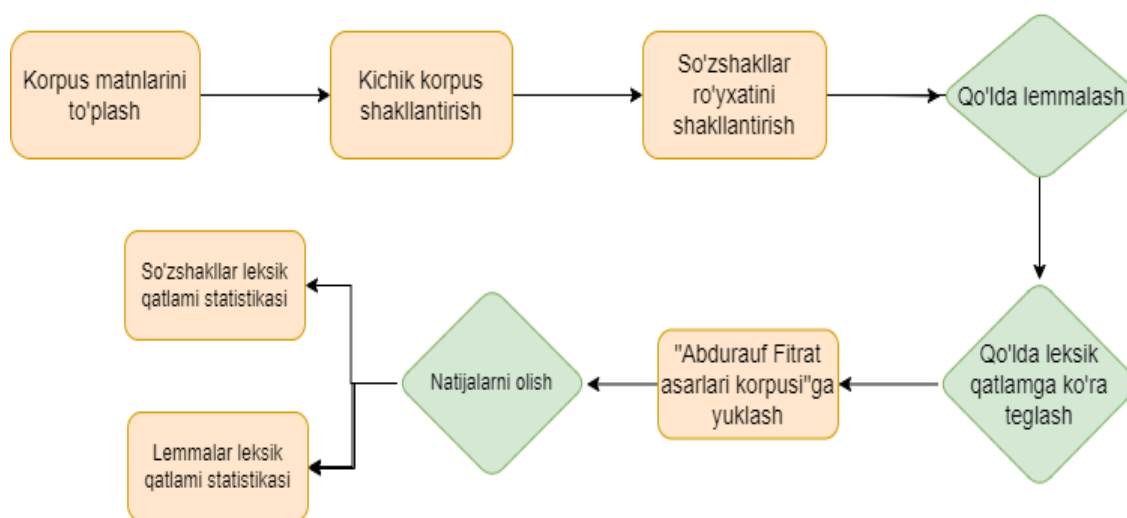
Ключевые слова: корпус Абдурауфа Фитрата, лемма, лексический слой, исконные слой, заимствованный слой, разметки.

Abdurauf Fitratning badiiy asarlari tilining leksik xususiyatlari Y.Saidov tomonidan o‘rganilgan. Fitratning badiiy asarlari tili o‘rganilgan bo‘lsa, bizning tadqiqotimizda muallifning barcha asarlari tilining leksikasi korpus yordamida o‘rganilib, statistik jihatdan tahlil qilindi. Leksik qatlamlari yarimavtomat tarzda teglandi. Abdurauf Fitrat asarlari korpusi yaratildi. Aniqlangan so‘zshakllar qo‘lda lemma shakliga keltirildi. So‘zshakllar ro‘yxatida so‘z bo‘lmagan birliklar, tinish belgilari, qo‘shimchalar ham mavjud. Shu kabi birliklar o‘chirilib, so‘zshakllar lemmatizatsiya qilindi. Ushbu jarayonda qo‘shma fe‘l ekani ma‘lum bo‘lgan, ammo dastur tomonidan ikkiga ajratilib qo‘yilgan so‘zlar aniqlanib, qo‘shma fe‘l shaklida lemmalandi. Bunga *xitob qilmoq, voyaga yetmoq, dod solmoq, ta‘qib qilmoq, taqozo etmoq* kabi fe‘llarni misol qilish mumkin. Atoqli otlar, jumladan, shaxs nomi, joy nomi, asar nomlari va boshqalar qo‘lda ajratib olindi. Shaxs nomlari ism, familiya, otasining ismi, taxallus shakllarida uchraydi. Ularning soni 707 tani tashkil etadi. *Muqimxon, Muso A‘lo Sha‘riy, Muso binni Nasir, Mutavakkil, N.S.Likoshin, Najib Osimbek, Najmiddin Roziy, Najoshiy* kabi atoqli otlarni bunga misol qilish mumkin. 178 ta joy nomlari aniqlandi. Ularning tarkibida, davlat nomlari, shahar va qishloq nomlari mavjud. Millatning nomi davlat nomi sifatida qo‘llanilgan o‘rinlar ham mavjud. Bir davlatning nomi bir necha variantda nomlangan holatlar ham kuzatildi: Farang, Farangiston, Faronsa, Fransiyo. Abdurauf Fitrat asarlari korpusi matnida jami 314 ta asar nomi keltirilgan. “Bahori ajam”, “Bahoriston”, “Bahr ul-abror”, “Bahr ul-asror”, “Bahrom-Gulandom”, “Binafsha”, “Bedil”, “Bobirnama” kabi asarlarni bunga misol qilish mumkin. Shuningdek, atab qo‘yilgan nomlardan to‘garak nomi, unvon nomi, yulduz nomi, jonivorlarning nomi, gazeta, jurnal va kuy nomlari qo‘llanilgan bo‘lib, ularning soni 46 tani tashkil etadi. Ushbu atoqli otlar qo‘lda lemmalar ichidan ajratib olindi.

So‘zshakllar asosida lemmalar qo‘lda lemmatizatsiya qilingach, 9157 ta lemmalar bazasi shakllantirildi. Ushbu lemmalar tarkibida dublet so‘zlar, muallif tomonidan turli shaklda

qo'llanilgan lemmalar ham mavjud. Bunga *yumshoqlik*, *yumshoqliq*; *yo'zg'uchi*, *yo'zguvchi*; *yomonlik*, *yomonliq*; *ahmoqlik*, *ahmoqliq*; *ambarlik*, *ambarliq* kabi so'zlarni misol qilish mumkin.

Shakllantirilgan lemmalarning leksik qatlami o'rganildi. Qo'llanilgan so'zlarning qatlamlarini aniqlashda Sh.Rahmatullayevning "O'zbek tilining etimologik lug'ati" [1], O'zbek tilining izohli lug'ati [2], E. Begmatovning "Hozirgi o'zbek tilining leksik qatlamlari" [3] kitoblaridan foydalanildi. Muallif ijodida qadimgi turkiy til va eski o'zbek adabiy tiliga oid so'zlar qo'llanilgani uchun Mahmud Qoshg'ariyning "Devoni lug'oti turk" [4], Alisher Navoiyning "Muhokamat ul-lug'atayn" [5] asarlariga ham murojaat qilindi. "Abdurauf Fitrat asarlari korpusi" [6] matnlarini leksik qatlamga ko'ra teglash algoritmi tasvirlangan (1-rasmga qarang).



1-rasm. "Abdurauf Fitrat korpusi"ning leksik qatlamlarini teglash algoritmi

Abdurauf Fitrat tilning sofligi masalasiga jiddiy qaragan va shu yo'lda maqolalar e'lon qilib, millat ziyolilarini turkiy so'zlardan foydalanishga, o'zlashma arabiy va forsiy so'zlardan imkon qadar foydalanmaslikka targ'ib qilgan. Muallifning asarlarida leksik qatlamning ifodalanishi, o'z va o'zlashgan qatlamga oid so'zlarni tahlil qilish, uning statistikasini aniqlash ahamiyatli. O'zbek tiliga so'zlar ko'pincha arab, fors, rus tillaridan o'zlashgan. O'zbek tilining ichki imkoniyati asosida so'zga qo'shimcha qo'shish orqali ham so'zlar yasaladi. O'zlashma so'zlarga o'zbek tilida mavjud bo'lgan yasovchi vosita qo'shish orqali ham so'zlar yasaladi va o'z qatlamdagi so'zlar soni ortadi. Shuningdek, qo'shimchalar ham o'zlashadi. Turkiy tillar xususiyatiga ko'ra, o'zbek tilida old qo'shimcha mavjud emas. Fors-tojik tilidan ba, be, bo, no, pur, ser kabi qo'shimchalar o'zlashgan. -iy, -viy, qo'shimchalari arab tilidan o'zlashgan. Bu kabi qo'shimchalar bilan yasalgan so'zni qatlamini aniqlashda ikki xil holat kuzatiladi:

1. Yasalishga asos bo'lgan qism o'zbek tilida mavjud bo'lmasa, yasalish fors- tojik tilida bo'lib, u so'z o'zlashgan so'z hisoblanadi: ma'naviy, boadab, ijtimoiy kabilar.

2. Yasalishga asos qism o'zlashgan so'z bo'lsa ham, u tilimizda qo'llansa, yasalgan so'z o'z qatlamga oid bo'ladi, ya'ni yasalish o'zbek tilida bo'lgan. Masalan, *betinim*, *notinch*, *kitobiy*.

Chetdan o'zlashgan so'zlar yordamida qo'shma fe'l yasalgan holatlar ham kuzatildi. Fe'lning yetakchi qismi o'z qatlamga oid bo'lmasa ham, uning ko'makchi qismi turkiy so'z bo'lgani va yasalish o'zbek tilida kuzatilgani uchun bunday so'zlar o'z qatlamga mansub deb belgilandi. Bunga *tarbiya qilmoq*, *arz qilmoq*, *afu etmoq*, *zoe qilmoq*, *madh etmoq* kabi so'zlarni misol qilish mumkin.

Chetdan o'zlashgan so'zlarga yasovchi qo'shimcha qo'shib yasalgan barcha so'zlar o'z qatlamga kiritildi. Masalan, *maktab* – arabcha so'z, *maktabdorlik* – turkiy so'z. Exel fayldagi lemmalar qaysi qatlamga mansub ekani o'rganilib, har bir so'z yoniga teglab chiqildi. Leksik qatlamga ko'ra teglangan ma'lumotlar "Abdurauf Fitrat korpusi"ga yuklandi va natijalar olindi.

Xulosa. Abdurauf Fitrat asarlari lemmalari korpus yordamida aniqlangan soʻzshakllar roʻyxati asosida qoʻlda lemmalashtirildi. Lemmalash jarayonida atoqli otlar aniqlanib, ularning alohida roʻyxati tuzildi va lemmalar bazasidan oʻchirildi. Shakllantirilgan lemmalar oʻz va oʻzlashma qatlamga koʻra qoʻlda teglab chiqildi.

Foydalanilgan adabiyotlar:

1. Раҳматуллаев Ш. Ўзбек тилининг этимологик луғати. Тошкент: Университет, 2003. – Б. 599.
2. Oʻzbek tilining izohli lugʻati: 80000 dan ortiq soʻz va soʻz birikmasi, VI jildlik. – Toshkent: Gʻafur Gʻulom, 2023. – 964 bet.
3. Бегматов Е. Ўзбек тилининг лексик қатламлари. – Тошкент: Фан, 1985. – Б. 199.
4. Қошғарий М. Девони луғоти турк, ЎзССР Фанлар академияси нашриёти, 1960. – Б. 498.
5. Navoiy A. Muhokamat ul-lugʻatayn. Toshkent: Tafakkur, 2014. – B. 137.
6. <https://fitratkorpusi.uz>

AVTOMATIK TARJIMADAGI MUAMMOLAR

Nurmonova Sadoqat

Andijon davlat universiteti
tilshunoslik kafedrası o`qituvchisi PhD

To`xtaboyeva Nasiba

Andijon davlat universiteti filologiya fakulteti talabasi

Annotatsiya: Ushbu tezisdagi kompyuter lingvistikasining sohalaridan biri bo`lgan avtomatik tarjima yo`nalishidagi bir qancha muammolar, xususan, Google tarjimon dasturi, katta hajmdagi matnlarni tahlil qiluvchi dastur va uning boshqa tarjimon dasturlardan farqi haqida ma`lumotlar keltirilgan.

Kalit so`zlar: kompyuter lingvistikasi, avtomatik tarjima, tarjimon dasturi, ibora, tarjima jarayoni.

Аннотация: В данной дипломной работе представлены некоторые проблемы в области автоматического перевода, который является одной из областей компьютерной лингвистики, в частности, сведения о программе-переводчике Google - программе, анализирующей большие объемы текстов, и ее отличиях от других программ-переводчиков.

Ключевые слова: компьютерная лингвистика, автоматический перевод, программа-переводчик, фраза, процесс перевода.

Abstract: This thesis presents information on several problems in the field of automatic translation, one of the areas of computational linguistics, in particular, the Google Translate program, a program that analyzes large volumes of text, and its differences from other translation programs.

Keywords: computational linguistics, automatic translation, translator software, phrase, translation process.

Texnika rivojlangan asrda yashayotgan ekanmiz, u kundalik hayotimiz bilan bevosita aloqaga kirishayotganini inkor etib bo`lmaydi. Xususan, uning boshqa fanlar va tilshunoslik bilan aloqasi kundan-kunga rivojlanib bormoqda. Jumladan, ma`lum bir tildagi asarni boshqa bir tilga tarjima qilish, umuman kichik bir matn, xoh so`z bo`lsin, kompyuter yordamida bu ishni amalga oshirish qulayroq ammo samaraliroq deb ayta olmaymiz.

Avvalambor, mashina tarjimasi tushunchasiga to`xtaladigan bo`lsak, mashina tarjimasi yoki avtomatik tarjima deyilganda bir tildagi matnni ikkinchi bir tilga EHM (kompyuter) vositasida, tez vaqt ichida tarjima qilish nazarda tutiladi.⁸⁵ Keyinchalik kompyuter lingvistikasi degan atama ham shundan kelib chiqqan.

Avtomatik tarjima kompyuter lingvistikasi fanining sohalaridan hisoblanadi. Aynan avtomatik tarjimada amalga oshirilgan ishni samarali deb ayta olmasligimizning sababi bir qancha. Misol uchun, oddiygina "Google" tarjimon dasturini olaylik. Agar ilmiy matnni tarjima qilmoqchi bo`lsak, tarjima jarayonida xatoliklar ko`p bo`lmaydi. Biroq badiiy matnni tarjima qiladigan bo`lsak, ko`plab xatoliklarga yo`l qo`yilganining guvohi bo`lamiz. Xususan, matnda muallif o`quvchiga yetkazib bermoqchi bo`lgan ma`nooni tarjima dasturi yetkazib bera olmasligi tabiiy. Sababi, badiiy matnda iboralardan, ayrim o`rinlarda shevaga xos so`zlardan ham foydalanilishi mumkin. Aytaylik, bu tarjima dasturining oldida turgan muammo. Uning yechimi qanday degan savol tug`iladi. Umuman olganda bu kabi muammolar haqida ko`p gapiriladi. Xususan, NLP (Natural language processing) ham shular jumlasidan.

⁸⁵ A.Rahimov. Kompyuter lingvistikasi asoslari. – T.: Akademnashr, 2011. – B.5

NLP matnni bir tildan ikkinchi bir tilga tarjima qiladigan, ogʻzaki buyruqlarga javob beradigan va katta hajmdagi matnni tez, hatto real vaqtda umumlashtiradigan kompyuter dasturlarini boshqaradi.⁸⁶ Aniqroq qilib aytganda NLP – bu kompyuterlarning inson tilini tushunishi. Shuningdek, u kompyuterdagi matn mazmunini anglaydi va Google Translate kabi xizmatlarni ham amalga oshiradi.

Yaʼni u tarjima jarayonida matnni bir tildan ikkinchi bir tilga tarjima qilayotganimizda, matnni tushunishga yordam beradi.

Aslida, bu dasturlar ham oʻzida bor maʼlumotlar orqali tarjimani amalga oshiradi. Yechim esa ana shu dasturlardagi bazani yana ham toʻldirish. Bundan tashqari, tarjimaning nechogʻlik sifatli tarjima qilinishi soʻz, soʻz birikmasi yoki iboralarning uslub jihatdan toʻgʻri tanlanganligiga ham bogʻliq. Shu bois matn qayta tahlil qilinib, segmentlash yoki tarjima jarayonidagi yoʻl qoʻyilgan xatolar tarjimon va tilshunoslar tomonidan tekshiriladi.⁸⁷

Avtomatik tarjimada samaradorlikning yuqori boʻlishiga tabiiy tillarning ayrim jihatlari – polisemiya (koʻp maʼnolilik), omonimlik, frazeologizmlar, idiomalar, paremlar, troplar toʻsqinlik qiladi.⁸⁸ Darhaqiqat, hozirgi kunda tarjima boʻyicha yuqorida sanab oʻtilgan muammolar ishning samarali boʻlishini orqaga suradi. Masalan, frazeologizmlarni birma-bir maʼnosini yozib, tarjima jarayonining bazasiga kiritilgan taqdirda ham, kontekstda notoʻgʻri maʼno anglatib qoʻyish ehtimol mavjud. Agar biz badiiy matnni boshqa bir tilga tarjima qilmoqchi boʻlsak, iboraga duch kelishimiz tabiiy. Xuddi mana shu holatda biz iborani yaxlitlab olishimizga toʻgʻri keladi. Yaʼni ibora maʼnosiga mos keluvchi muqobil kontekst bilan berilishi shart. Bitta gap misolida koʻraylik. “Salima tomdan tarasha tushganday turgan joyida qolib ketdi.” Bu gapni ingliz tiliga tarjima qilamiz. Google translate orqali tarjima qilinganda, “*Salima remained where she was standing as if she had fallen from the roof*” tarjimasi kelib chiqdi. Gapdagi “tomdan yiqilgandek” deb tarjima qildi. Aslida, bu ibora “toʻsatdan”, “birdan” degan maʼnoni beradi.

Shuningdek, mashina tarjimasi dasturlari dunyo boʻylab keng tarqalgan. Bir qancha olimlar va tadqiqotchilar, odamlar tarjima dasturlaridan asosan qanday tarjimalarni amalga oshirishiga qiziqishgan. Xususan, Dyuk universitetida 900 dan ortiq bakalavr talabalari oʻrtasida soʻrovnoma oʻtkazilgan. Ishtirokchi talabalarning koʻpchiligi tizimga alohida soʻzlarni kiritish orqali Google tarjimonidan onlayn lugʻatga oʻxshash tarzda foydalanilishi haqida xabar beradi.⁸⁹ Toʻgʻri, agar dasturdan onlayn lugʻat tarzida foydalanilsa, undagi kamchiliklar unchalik ham sezilmasligi mumkin. Lekin iboralar, yuqorida aytilganidek, muammoli masala.

Yandex Translate va Google Translate tarjima dasturlari oʻrtasidagi farqlar iboralarni tarjima qilishdagi yondashuvlar, dasturiy algoritmlar asosida tahlil qilinadi. Masalan Google Translate, qoidalarga asoslangan holda tarjimani amalga oshiradi va tarjima natijalari asosan matnlar asosidagi oʻxshashliklarga asoslanadi. Yandex translate xizmati esa statik metodga asoslangan holda tarjima jarayonida morfologik tahlilni amalga oshiradi.⁹⁰ Masalan, “*Ogʻzi qulogʻida*” iborasi, Google Translate dasturida “*his mouth is in his ear*” tarzida, Yandex Translate dasturida esa “*In the ear of the mouth*” tarzida tarjima qiladi. Biroq, ChatGPT esa “*Grinning from ear to ear*” yoki “*Beam with joy*” tarzida tarjima qilmoqda. Vaholangki, yuqoridagi sanalgan ikki tarjimon dasturiga qaraganda ChatGPTning tarjimasi nisbatan aniqroq. Sababi ChatGPT, OpenAI tomonidan yaratilgan sunʼiy intellekt modeli boʻlib, uning asosiy vazifalaridan biri tabiiy tilni tushunish. U katta til modellari (Large Language Model) asosida ishlaydi.

Birinchidan, LLM (Large Language Model) resurslarni tejaydigan tarzda tarjima qilish qobiliyatiga ega boʻlishi va hatto manbasiz tillarda ham oʻrtacha tarjima yaratishi mumkin.

⁸⁶ M.Adilova. NLP vazifalari va yoʻnalishlari// “Kompyuter lingvistikasi: muammolar, yechim, istiqbollar”.Xalqaro ilmiy-amaliy konferensiya,2023.

⁸⁷ N.Abdurahmonova. Kompyuter lingvistikasi. –T.: Nodirabegim nashriyoti,2021. –B.36

⁸⁸ Z.Xolmonova. Kompyuter lingvistikasi. –T.: Asian Book House nashriyoti, 2020. –B.140

⁸⁹ Per Urlaub and Eva Dessein. Machine Translation and foreign language education //Global Language, Massachusetts institute of Technology, Cambridge, MA United States

⁹⁰ S.Sharipova, N.Sidikova. Iboralarni avtomatik tarjima qilish bilan bogʻliq ayrim muammolar//International Conference:The development history and significance of intercultural dialogue

Ikkinchidan, o`zaro tillardagi misollar bir xil til juftliklaridagi misollarga kam manbali tarjima uchun yaxshiroq vazifa bo`yicha ko`rsatmalar berishi mumkin.<sup>91</sup> Xulosa o`rnida aytganda, LLM katta hajmdagi matnlarni tahlil qilib uning konteksdagi tushunishini ham o`rganadi. Google Translate va Yandex Translate mashinaga yuklangan matnlarga asoslanib o`rgangani va ma`lumotlar bazasining cheklanganligi uchun ham tarjimada bir qancha xatoliklarga yo`l qo`yishi mumkin. Shuning uchun ham hozirgi kunda sun`iy intellektga asoslangan tarjima vositalari nisbatan mukammalroq.

Demak, biz badiiy matnni tarjima qilayotganimizda iboraning o`rniga uning bir so`zga teng muqobilini qo`yishimizga to`g`ri keladi. Bundan tashqari o`zbek tilida birgina so`z kontekstda har xil ma`no anglatishi mumkinligini ham hisobga olish zarur. Vaholangki, boy tilimizda bunday rang-barang iboralar ko`p uchraydi. Ularni kundalik hayotimizda ham ko`p uchratamiz. Mana shu iboralarimizdan boshqa tilda ham unumli va erkin foydalana olishimiz uchun tilshunos va dasturchilardan mashaqqatli mehnat talab qiladi.

Foydalanilgan adabiyotlar:

1. Rahimov. Kompyuter lingvistikasi asoslari. –T.:Akademnashr. -158 bet
2. M.Adilova .NLP vazifalari va yo`nalishlari.//”Kompyuter lingvistikasi: muammolar, yechim, istiqbollar”. Xalqaro ilmiy-amaliy konferensiya.
3. N. Abdurahmonova. Kompyuter lingvistikasi. –T.:Nodirabegim. -394 bet
4. Z.Xolmonova. Kompyuter lingvistikasi. –T.:Asian Book House. -246 bet
5. Per Urlaub and Eva Dessein. Machine Translation and foreign language education. Global language Massachusetts Institute of technology, Cambridge, MA United States
6. S. Sharipova N. Sidikova. Iboralarni avtomatik tarjima qilish bilan bog`liq ayrim muammolar.//International conference: The development history and significance of intercultural dialogue.
7. Wenhao Zhu, Hongyi Liu, Qingxiu Dong, Jingjing Xu Shujian Huang, Lingpeng Kong, Jiajun Chen, Lei Li. Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis.

⁹¹ Wenhao Zhu, Hongyi Liu, Qingxiu Dong, Jingjing Xu, Shujian Huang, Lingpeng Kong, Jiajun Chen, Lei Li // Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis.

O'ZBEK TILIDA KOLLOKATSIYALARGA DOIR YONDASHUVLAR

Yuldasheva Ruxshona

O'zbekiston Milliy universiteti

Kompyuter lingvistikasi mutaxassisligi magistranti

Annotatsiya: Kollokatsiyalar muhim lingvistik birlik bo'lib, ular so'zlarning bir-biri bilan doimiy bog'lanishini ta'minlaydi va mazmun jihatidan aniq, tabiiy iboralarni shakllantiradi. Grammatik nuqtai nazardan, kollokatsiyalar morfologik va sintaktik tuzilmalarga asoslangan bo'lib, ular tilning semantik va stilistik jihatlariga ham bevosita ta'sir ko'rsatadi.

Ushbu maqolada kollokatsiyalarning grammatik xususiyatlari tahlil qilinadi, jumladan, so'z turkumlari o'rtasidagi bog'lanishlar, sintaktik birliklarning tuzilishi va grammatik moslashuv masalalari yoritiladi. Shuningdek, zamonaviy o'zbek tilida kollokatsiyalarni o'rganishning ahamiyati va ularning tarjima jarayonidagi murakkab jihatlar muhokama qilinadi.

Kalit so'zlar: kollokatsiya, leksik birlik, grammatik uyg'unlik, sintaktik bog'lanish, o'zbek tili, semantik moslashuv, morfologik xususiyat, tarjima muammolari.

Abstract: Collocations are an important linguistic unit that ensures the constant connection of words with each other and forms clear, natural expressions in terms of content. From a grammatical point of view, collocations are based on morphological and syntactic structures, which also directly affect the semantic and stylistic aspects of the language.

This article analyzes the grammatical features of collocations, including the connections between word groups, the structure of syntactic units, and grammatical adaptation. It also discusses the importance of studying collocations in modern Uzbek and their complex aspects in the translation process.

Keywords: collocation, lexical unit, grammatical harmony, syntactic connection, Uzbek language, semantic adaptation, morphological feature, translation problems.

Аннотация: Коллокации — важная языковая единица, обеспечивающая постоянную связь слов друг с другом и образующая осмысленные, естественные выражения. С грамматической точки зрения словосочетания основаны на морфологических и синтаксических структурах, которые также напрямую влияют на семантические и стилистические аспекты языка.

В статье анализируются грамматические свойства словосочетаний, включая связи между классами слов, структуру синтаксических единиц и грамматическую адаптацию. Также обсуждается важность изучения устойчивых словосочетаний в современном узбекском языке и их сложности в процессе перевода.

Ключевые слова: словосочетание, лексическая единица, грамматическая связность, синтаксическая связь, узбекский язык, семантическая адаптация, морфологическая особенность, проблемы перевода.

Kollokatsiyalar tilda tabiiy va aniq ifodalash vositasi bo'lib, ular aniq ma'no hosil qiluvchi mustahkam birikmalarni tashkil etadi. Har bir tilning o'ziga xos grammatik va semantik qonuniyatlari mavjud bo'lgani sababli, kollokatsiyalar so'zlarning grammatik shakllari va ularning o'zaro bog'lanish qoidalariga qat'iy rioya qiladi.

Kollokatsiyalarni o'rganish nafaqat leksikologiya, balki morfologiya va sintaksis nuqtai nazaridan ham muhim. Chunki kollokatsiyalarni tashkil etuvchi so'zlar o'zaro grammatik uyg'unlik asosida bog'lanadi va semantik moslashuv natijasida barqaror iboralarni hosil qiladi.

Mazkur maqolada kollokatsiyalarning grammatik tuzilishi, ularning tildagi o'rni, morfologik va sintaktik xususiyatlari keng tahlil qilinadi.

Tadqiqot metodologiyasi va empirik tahlil

Kollokatsiyalar morfologik birliklarning o‘zaro bog‘lanishiga asoslanadi. Kollokatsiyalarni tashkil qiluvchi birliklar so‘z turkumlariga ko‘ra turlicha bo‘lishi mumkin. Quyida ularning asosiy turlari keltirilgan:

Kollokatsiya turi	Tuzilish shakli	Misollar
Fe‘l + Ot	Fe‘l va otning birikishi	qaror qabul qilish, yordam berish
Ot + Ot	Bir ot ikkinchi otga bog‘lanadi	ilmiy ish, mehnat taqsimoti
Sifat + Ot	Sifat va ot birikmasi	qattiq sovuq, baland ovoz
Fe‘l + Fe‘l	Ikki fe‘l birikadi	e‘tibor qaratmoq, hisobga olmoq
Raqam + Ot	Son va ot kombinatsiyasi	bir necha yil, o‘n kishi

Ushbu misollar shuni ko‘rsatadiki, kollokatsiyalar morfologik tuzilish jihatidan turli kombinatsiyalarni hosil qilishi mumkin, bu esa ularning grammatik moslashuvchanligini ta‘minlaydi.

Kollokatsiyalar sintaktik jihatdan mustahkam bog‘lanish asosida shakllanadi. Kollokatsiyalar sintaktik tuzilishga ko‘ra quyidagi xususiyatlarga ega:

Mustahkam bog‘lanish – so‘zlar o‘rtasidagi bog‘liqlik kuchli bo‘lib, ular o‘zaro almashtirib bo‘lmaydi. Masalan, “qaror qabul qilish” iborasini “qaror olish” deb o‘zgartirib bo‘lmaydi.

Qoidali tartib – o‘zbek tilida kollokatsiyalar o‘z tartibini saqlaydi. Masalan, “katta imkoniyat” deb aytiladi, “imkoniyat katta” shakli esa kam qo‘llaniladi.

Bo‘g‘in soni va urg‘u – kollokatsiyalar talaffuz jihatdan ma‘lum ritm va urg‘uga ega bo‘ladi.

Sintaktik bog‘lanishning mustahkamligi kollokatsiyalarni erkin so‘z birikmalaridan farqlaydi va o‘zbek tilidagi tabiiy nutqning ajralmas qismi hisoblanadi.

Kollokatsiyalar tarjima jarayonida murakkablik tug‘dirishi mumkin, chunki ularning ekvivalentini topish faqat so‘zma-so‘z tarjima orqali emas, balki semantik moslashuv orqali amalga oshiriladi.

Kollokatsiya	Ingliz tilidagi ekvivalent	Tarjima muammolari
Qaror qabul qilish	Make a decision	Fe‘l to‘g‘ridan-to‘g‘ri tarjima qilinmaydi
Yurakdan gapirish	Speak from the heart	Idiomatik moslashuv talab etiladi
Yaxshi niyat bilan	With good intentions	Kontekstga qarab moslashadi

Ushbu jadvaldan ko‘rinib turibdiki, tarjima jarayonida kollokatsiyalar o‘ziga xos yondashuvni talab qiladi va ularni bevosita tarjima qilish har doim ham ma‘no jihatdan mos kelavermaydi.

Kollokatsiyalar o‘zbek tilida nafaqat so‘zlar o‘rtasidagi grammatik bog‘lanishni ta‘minlaydi, balki ularning ma‘no jihatidan ham muhim rol o‘ynaydi. Kollokatsiyalarni o‘rganish orqali tilning tabiiy ifodalanish qoidalari va nutqning aniq va tushunarli bo‘lishi ko‘rsatiladi. Ularning tahlili orqali o‘zbek tilining semantik tuzilmasi va nutqning dinamikasini tushunish mumkin.

Kollokatsiyalar tilni o‘rganishda muhim ahamiyatga ega, chunki ular til o‘rganuvchisi uchun tabiiy ifodalarni o‘rganish imkonini beradi. Kollokatsiyalarni o‘rganish o‘quvchilarga so‘zlarni to‘g‘ri va mos ravishda ishlatishga yordam beradi. O‘zbek tilida bunday birikmalar so‘zning o‘ziga xos qo‘llanish doirasini belgilaydi va ularning turg‘unligiga olib keladi. Misol uchun, “qiziqish bildirish” va “qiziqish uyg‘otish” iboralari bir-biriga yaqin bo‘lsa-da, ular turli kontekstlarda qo‘llanadi. Bu nuqtai nazardan, kollokatsiyalar nafaqat grammatikaning muhim qismidir, balki til o‘rganishda ham muhim vosita hisoblanadi.

Kollokatsiyalar stilistik jihatdan ham katta ahamiyatga ega. Kollokatsiyalarni turli kontekstlarda ishlatish xususiy uslubni yaratadi, shuningdek, nutqning tabiiyligini ta‘minlaydi.

Kollokatsiyalarni o'zbek tilida ishlatish, o'z navbatida, ma'lum bir nutq uslubini yaratadi. Misol uchun, ilmiy nutqda "yo'l-yo'riq berish" kabi ifodalar keng qo'llaniladi, adabiy nutqda esa "yo'l ko'rsatish" ko'proq ishlatiladi. Har bir kollokatsiya o'zining uslubiy izchilligini saqlaydi, shuningdek, uni to'g'ri tanlash nutqning yo'nalishini aniqlaydi.

Shu o'rinda, kollokatsiyalarni o'zbek tilida adabiyot, ommaviy axborot vositalari va rasmiy hujjatlar kabi turli sohalarida ishlatish stilistik jihatdan farq qiladi. Rasmiy so'z birikmalari ko'pincha yuridik, siyosiy va iqtisodiy mavzularda keng qo'llanadi, masalan, "qaror qabul qilish", "moliyaviy holatni tahlil qilish", yoki "davlat siyosatini amalga oshirish" kabi ifodalar.

Kollokatsiyalarni o'rganishda semantik moslashuvning ahamiyati ham katta. Kollokatsiyalar ko'pincha ma'lum bir semantik yondashuvni talab qiladi. Ular bir-biriga mantiqan, semantik jihatdan uyg'un keladigan so'zlardan tashkil topgan bo'lib, ularni o'zaro almashtirish mumkin emas. Misol uchun, "yangi imkoniyat yaratish" va "yangi imkoniyat ochish" iboralari o'zaro almashtirilsa ham, ular turli semantik yuksalishlarga ega bo'ladi. "Yangi imkoniyat yaratish" iborasi odatda biror yangilikni joriy qilishni bildiradi, "yangi imkoniyat ochish" esa avvalgi imkoniyatlar ustida yangiliklar kiritishni anglatadi.

Shuningdek, kollokatsiyalar turli tillar o'rtasida tarjima qilinishi jarayonida semantik o'zgarishlar yuzaga kelishi mumkin. Misol uchun, "qattiq to'xtatish" ifodasi ingliz tiliga "stop abruptly" tarzida tarjima qilinsa-da, o'zbek tilida bu ibora ko'proq kuchli to'xtash yoki ta'sirli to'xtash ma'nosini anglatadi.

Kollokatsiyalarni o'rganish o'zbek tilining grammatikasini chuqurroq anglashga yordam beradi. Kollokatsiyalarni morfologik va sintaktik jihatdan tahlil qilish orqali tilshunoslar tilning yanada rivojlanish yo'llarini belgilashlari mumkin. Ularning yordamida tilning struktura tizimini yanada mukammalroq o'rganish va tilni to'g'ri ishlatish qoidalarini aniqlash mumkin. Kollokatsiyalarni tahlil qilish nafaqat tilning o'ziga xosligini, balki tilning taraqqiyotini va zamonaviy ehtiyojlarni ham ko'rsatadi.

Kollokatsiyalarni tarjima qilishda yuzaga keladigan muammolar morfologik va sintaktik differentsialarni aniqlashda yordam beradi. O'zbek tili va boshqa tillar o'rtasidagi kollokatsiyalarni to'g'ri tarjima qilishda semantik moslashuv zarur. Tarjimonlar kollokatsiyalarni leksik va grammatik jihatdan to'g'ri tanlash orqali tilni nafaqat so'zma-so'z, balki semantik tarzda ham o'zgartirishga majburdirlar. Boshqa tillarga nisbatan o'zbek kollokatsiyalarining maxsus ifodalanish uslubi ham tarjimaga oid masalalarni qiyinlashtiradi. Shuningdek, turli tillardagi kollokatsiyalarni taqqoslash tilshunoslarga tilning ixtisoslashgan jihatlarini aniqlash imkonini beradi.

Kollokatsiyalar til egasining bilimi va tajribasiga asoslangan holda shakllanadi. Bu ularning kognitiv tilshunoslik bilan bog'liqligini ko'rsatadi. Bunda asosiy jihatlari:

Tushuncha xaritalari – Kollokatsiyalar so'zlar o'rtasida kuchli semantik bog'lanish hosil qiladi. Masalan, "qaror qabul qilish" iborasi faqat "qabul" fe'li bilan ishlatilishi odatiy bo'lib qolgan.

Assotsiativ fikrlash – Kollokatsiyalar ko'pincha bir-biriga bog'liq tushunchalar orqali yodda saqlanadi. Masalan, "baland ovoz" va "qattiq shamol" iboralari semantik jihatdan yaqin bo'lsa ham, ularning kollokativ kombinatsiyasi turlicha bo'ladi.

Til tizimidagi mustahkam o'rin – Kollokatsiyalar so'zlarning stilistik va kontekstual uyg'unligini saqlab turadi. Masalan, "yaxshi imkoniyat" iborasi mavjud bo'lib, "yaxshi qaror" kollokatsiyasi kam ishlatiladi.

Kollokatsiyalar psixolingvistika nuqtai nazaridan til egasi tomonidan tez va avtomatik tarzda tanib olinadigan iboralar hisoblanadi. Bu jarayon quyidagicha amalga oshadi:

Tezkor idrok va qayta ishlash – Kollokatsiyalar ong osti darajasida tezkor tahlil qilinadi. Masalan, o'zbek tilida "dasturxon yozish" iborasi darhol ma'no hosil qilsa, "dasturxon ochish" iborasi g'alati tuyuladi.

Oson yodda saqlash – O‘quvchilar yangi so‘zlarni mustahkam bog‘langan kollokatsiyalar shaklida o‘rgansa, ularni eslab qolish osonlashadi. Masalan, “yo‘l harakati qoidalari” iborasi “yo‘l qoidalari”dan ko‘ra aniqroq esda qoladi.

Til o‘rganish jarayonidagi ahamiyati – O‘zbek tilini ikkinchi til sifatida o‘rganuvchilar uchun kollokatsiyalar mustahkam yodlash va tabiiy nutq shakllantirishda muhim ahamiyat kasb etadi.

Kollokatsiyalar tarjimada katta qiyinchiliklar tug‘dirishi mumkin, chunki har bir til o‘ziga xos kollokatsion birliklarga ega. Masalan:

Kollokatsiya	Inglizcha ekvivalenti	To‘g‘ridan-to‘g‘ri tarjima
Qaror qabul qilish	Make a decision	Xato: Accept a decision
Tinchlik o‘rnatish	Establish peace	Xato: Make peace
Xato qilish	Make a mistake	Xato: Do a mistake

Bu misollardan ko‘rinib turibdiki, kollokatsiyalarni tarjima qilishda semantik moslashuv zarur, chunki har bir tilda so‘zlarning tabiiy bog‘lanish tamoyillari har xil.

Xulosa va munozara

Kollokatsiyalar grammatik, sintaktik va semantik jihatlardan turlicha tuzilgan bo‘lib, ular tilni aniq, tabiiy va murakkab ifodalash uchun zarur bo‘lgan so‘z birikmalarini tashkil etadi. Kollokatsiyalarni o‘rganish nafaqat tilni o‘rganishning samarali vositasi, balki tilshunoslikda yangi yondashuvlarni yaratish imkoniyatidir.

Kollokatsiyalar morfologik va sintaktik jihatdan mustahkam bo‘lib, ular tilning semantik mohiyatini ifodalashda ham katta ahamiyatga ega. Tarjima jarayonida kollokatsiyalarni semantik moslashuv asosida o‘rganish va xarakterli grammatik tuzilmalarga e‘tibor qaratish zarur. Kollokatsiyalarni chuqurroq o‘rganish til o‘rganishda samarali yondashuvni ta‘minlashga yordam beradi va o‘zbek tili rivojida muhim rol o‘ynaydi.

Kollokatsiyalarni to‘liq lug‘atlarni tuzish va ularni onlayn platformalarda taqdim etish zarur.

Tarjimonlar uchun kollokatsiyalarning to‘g‘ri tarjimasi bo‘yicha maxsus qo‘llanmalar ishlab chiqish lozim.

Kollokatsiyalarni grammatik va stilistik jihatlarini o‘rganish uchun maxsus ilmiy tadqiqotlar olib borish kerak.

Kollokatsiyalarni o‘rganish o‘zbek tilining rivojlanishida, tarjimonda va boshqa tillar bilan solishtirilganda yanada aniqroq tushunishni ta‘minlaydi.

Adabiyotlar ro‘yxati:

1. G‘ulomov A. Hozirgi o‘zbek adabiy tili. – Toshkent: O‘zbekiston Milliy Ensiklopediyasi, 2015.
2. Yo‘ldoshev S. O‘zbek tilining grammatik qurilishi. – Toshkent: Fan va texnologiya, 2017.
3. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
4. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
5. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
6. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus. uz. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.
7. Karimov B. O‘zbek tilshunosligi: Morfologiya va sintaksis. – Toshkent: Universitet, 2019.

8. Rasulov M. Tilshunoslik asoslari va o'zbek tilining grammatik tuzilishi. – Samarqand: Zarafshon, 2016.
9. O'zbek tilining izohli lug'ati. – Toshkent: O'zbekiston Fanlar Akademiyasi, 2020.

DIAXRON KORPUS YARATISHNING NAZARIY VA AMALIY AHAMIYATI

Yangibayeva Surayyo

ToshDO‘TAU, Kompyuter lingvistikasi mutaxassisligi

I bosqich magistranti

E-mail: surayyoyangibayeva4@gmail.com

Annotatsiya: Ushbu maqolada mustaqillik davri nashrlariga asoslangan diaxron korpus yaratishning nazariy va amaliy ahamiyati tahlil qilinadi. Mustaqillikdan keyin o‘zbek tilining rasmiy maqomga ega bo‘lishi, yangi so‘zlar va terminlarning kirib kelishi, grammatik va sintaktik o‘zgarishlar, matnlar uslubining yangilanishi kabi jarayonlar diaxron korpus orqali tizimli ravishda o‘rganilishi mumkin. Tadqiqotda korpus lingvistikasining asosiy tamoyillari, tarixiy matnlarni raqamlashtirish va belgilash texnologiyalari hamda o‘zbek tilining zamonaviy rivojlanish tendensiyalarini aniqlash imkoniyatlari ko‘rib chiqildi.

Kalit so‘zlar: diaxron korpus, mustaqillik davri nashrlari, kodifikatsiya, transformatsiyalar, lingvistik ta‘minot, korpus lingvistikasi.

Annotation: This article analyzes the theoretical and practical significance of constructing a diachronic corpus based on publications from the independence period. The transition of the Uzbek language to an official status after independence, the introduction of new words and terms, grammatical and syntactic changes, and the evolution of textual styles can be systematically studied through a diachronic corpus. The study examines the fundamental principles of corpus linguistics, technologies for digitizing and annotating historical texts, and the identification of contemporary trends in the development of the Uzbek language.

Keywords: diachronic corpus, independence period, codification, transformations, linguistic resources, corpus linguistics, digitization.

Аннотация: В данной статье анализируется теоретическая и практическая значимость создания диахронного корпуса, основанного на изданиях периода независимости. Системное изучение таких процессов, как приобретение узбекским языком официального статуса после независимости, появление новых слов и терминов, грамматические и синтаксические изменения, обновление стилистики текстов, возможно с использованием диахронного корпуса. В исследовании рассматриваются основные принципы корпусной лингвистики, технологии оцифровки и разметки исторических текстов, а также возможности выявления современных тенденций развития узбекского языка.

Ключевые слова: диахронный корпус, период независимости, кодификация, трансформации, лингвистическое обеспечение, корпусная лингвистика, оцифровка.

Kirish

Mustaqillikdan keyin o‘zbek tili davlat tili sifatida rivojlanishi yangi ilmiy tadqiqot yo‘nalishlarini kun tartibiga chiqardi. Ayniqsa, mustaqillik davrida chop etilgan matnlarni lingvistik tahlil qilish orqali til taraqqiyotining asosiy bosqichlarini o‘rganish muhim masalaga aylandi. Bunday tadqiqotlarni amalga oshirish uchun diaxron korpus yaratish zarurati paydo bo‘ldi. Diaxron korpuslar nafaqat o‘zbek tilining leksik va grammatik o‘zgarishlarini kuzatish, balki til tarixi va uning turli kontekstlardagi rivojlanish tendensiyalarini o‘rganish imkonini beradi. Shuningdek, u til o‘zgarishlarini matematik modellashtirish, sun‘iy intellekt texnologiyalari orqali semantik tahlil qilish va avtomatlashtirilgan tarjima tizimlari uchun asos sifatida xizmat qilishi mumkin. Ushbu tadqiqotda diaxron korpusning ilmiy va amaliy ahamiyati, uning tarkibiy qismlari, texnologik yondashuvlari va istiqboldagi rivojlanish yo‘nalishlari yoritiladi.

Asosiy qism

1. Mustaqillik davri diaxron korpusining nazariy asoslari

Diaxron korpus tilning turli davrlardagi rivojlanishini kuzatish imkonini beradigan tuzilma bo'lib, u lingvistik tadqiqotlarda muhim manba hisoblanadi [Sinclair J., 1991]. Mustaqillik davrida o'zbek tilida yuz bergan o'zgarishlarni chuqur tahlil qilish uchun bu davrga oid matnlarni jamlab, ularni lingvistik jihatdan belgilash talab etiladi. Diaxron tahlil tilning rivojlanish bosqichlarini aniqlash, leksik va sintaktik transformatsiyalarni kuzatish va lingvistik o'zgarishlarni statistik tahlil qilish imkonini beradi. Ushbu jarayonda quyidagi nazariy jihatlar asosiy o'rin tutadi.

Lingvistik kodifikatsiya: Mustaqillik yillarida o'zbek tiliga yangi so'zlar va atamalar kirib keldi, shuningdek, ayrim so'zlarning semantikasi o'zgardi. Bu jarayon davlat tilining rasmiy maqomini mustahkamlashga va ilmiy-texnik sohalarda terminologik barqarorlikni ta'minlashga xizmat qildi. O'zbek tilining normativ lug'atlari va grammatik qoidalariga doimiy o'zgartirishlar kiritilib, til tizimining yanada rivojlanishiga yo'l ochildi [Akademik N., 2020].

Leksik o'zgarishlar: Normativ lug'atlar va grammatik qoidalar ishlab chiqildi, bu esa diaxron tadqiqotlar uchun muhim manbadir. O'zbek tilida yangi terminlar, xalqaro so'zlar va mintaqaviy dialektlardan kelib chiqqan birliklar leksik o'zgarishlarning asosiy omillari sifatida namoyon bo'lmoqda. Shuningdek, globalizatsiya va texnologik taraqqiyot natijasida tildagi neologizmlar soni ortib bormoqda [Husainova G., 2018].

Uslubiy transformatsiyalar: Davlat va ilmiy nashrlarda, ommaviy axborot vositalarida rasmiy, ilmiy va publitsistik uslublarning o'zgarishi kuzatilmoqda. Mustaqillikdan so'ng rasmiy uslub yanada aniq va huquqiy-me'yoriy xarakter kasb etdi, ilmiy uslub esa xalqaro ilmiy tendensiyalar bilan uyg'unlashib, texnik va akademik terminlar bilan boyidi. Publitsistik uslubda esa erkinlik kuchayib, ijtimoiy va madaniy mavzularga e'tibor oshdi [Karimov I., 1997].

2. Mustaqillik davri diaxron korpusining amaliy ahamiyati

Diaxron korpus yaratish faqatgina nazariy tadqiqotlar uchun emas, balki amaliy jihatdan ham katta ahamiyatga ega. Ushbu korpus lingvistik tadqiqotlar, tarjimashunoslik, sun'iy intellekt asosidagi til texnologiyalari va madaniy merosni saqlash kabi turli yo'nalishlarda keng qo'llanilishi mumkin. Bundan tashqari, diaxron tahlil orqali til o'zgarishlarining statistik va semantik jihatlari chuqur o'rganiladi. Lingvistik birliklarning xususiyatlari, ularning nutq tarkibidagi xossalarga oid minglab misollarning konkordansda aniqlanishi korpusning birlamchi xususiyati sanalib, har qanday tilshunos lisoniy tadqiqotlarda korpus lingvistikasining yutuqlaridan foydalansa, tezkorlik, statistik aniqlik va empirik tahlilga yetarli misollarga osonroq yetisha oladi [Ataboyev N.B., 2024]. Bu kabi korpuslar faqat tilshunoslikda emas balki, turli sohalarda ham yangi tadqiqotlar eshigini ochadi. Ushbu korpus yordamida quyidagi sohalarda yangi imkoniyatlar ochiladi:

Tilshunoslik: O'zbek tilining so'nggi 33 yildagi rivojlanish dinamikasini o'rganish imkoniyati yaratiladi.

Tarjimashunoslik: Yangi lug'aviy birliklar va terminlarning tadqiqi orqali tarjimalarning sifatini oshirishga xizmat qiladi.

Axborot texnologiyalari: Sun'iy intellekt asosida til modellarini rivojlantirish va tabiiy tilni qayta ishlash (NLP) sohasida foydali bo'lishi mumkin [Hasanova M., 2021].

Madaniy merosni saqlash: Mustaqillik davri matnlarini saqlash va kelajak avlodlarga yetkazish imkoniyati yaratiladi.

Mana shu imkoniyatlarni hamda tilshunoslik uchun diaxron korpusning qanchalik muhim ekanligini inobatga olgan holda o'zbek tilining mustaqillik davri nashrlariga (asosan badiiy asarlar) asoslangan 1991-yildan 2024-yillar oralig'ini o'zida jamlagan diaxron korpusni yaratish ishlari olib borilmoqda.

3. Korpusni yaratish bosqichlari va texnologik yondashuvlar

Diaxron korpusni yaratish jarayoni quyidagi muhim bosqichlardan iborat bo'lib, har biri tilshunoslik va axborot texnologiyalarining ilg'or usullaridan foydalangan holda amalga oshiriladi:

1-jadval. Diaxron korpus yaratish jarayoni bosqichlari

Matn to'plash:	Mustaqillik yillarida chop etilgan gazeta, jurnal, ilmiy maqolalar va rasmiy hujjatlar yig'iladi.
Matnlarni raqamlashtirish:	OCR (optik belgilarni tanish) texnologiyalari yordamida bosma matnlar raqamlashtiriladi.
Leksik belgilash:	TreeTagger yoki UDPipe kabi dasturlar yordamida matnlar grammatik jihatdan belgilab chiqiladi.
Leksik normalizatsiya:	Tarixiy va zamonaviy til o'rtasidagi farqlar bartaraf etilib, yagona standartga keltiriladi.

Tanlangan yondashuvdan qat'iy nazar, korpuslarning veb-saytida dastlab **lemmatizatsiyani** amalga oshirish uchun ushbu qadamlarni bosib o'tish muhimdir [Xusainova Z., 2023]. Bundan tashqari, zamonaviy sun'iy intellekt vositalaridan foydalangan holda, matnlarning avtomatik semantik tahlili va kontekstual analiz tizimlari ham yo'lga qo'yilishi mumkin. Bu jarayon tildagi ma'no o'zgarishlarini aniqlash, sinonimiya va antonimiya tahlili hamda tilning dinamik rivojlanishini prognoz qilish imkonini beradi. Yana diaxron korpusni yaratish jarayonida NLP ham muhim vazifalarni bajaradi. Xususan, matnni tasniflash vazifasi. Bu matnning mazmuniga ko'ra ma'lum toifalar to'plamiga ajratish vazifasidir. Matnni tasniflash bugungi kundagi NLPdagi eng mashhur vazifa bo'lib, elektron pochta spamlarini aniqlashdan tortib hissiyotlarni tahlil qilishgacha bo'lgan turli xil vositalarda qo'llanishi fikrimiz isbotidir[Elov B.B., Hamroyeva Sh.M., Xusainova Z.Y., 2022].

Xulosa

Mustaqillik davriga oid diaxron korpus yaratish o'zbek tilining rivojlanish jarayonlarini chuqur tadqiq etish uchun muhim ahamiyatga ega. Bu korpus tilshunoslik, tarjimashunoslik, axborot texnologiyalari va madaniy merosni saqlash kabi sohalarida katta foyda keltiradi. Korpus lingvistikasi tamoyillariga asoslangan holda tarixiy matnlarni jamlash, ularni lingvistik jihatdan belgilash va analiz qilish orqali mustaqillik yillarida yuz bergan o'zgarishlarni ilmiy asosda o'rganish imkoniyati yaratiladi. Bundan tashqari, diaxron korpus sun'iy intellekt va ma'lumotlarni avtomatik tahlil qilish tizimlari bilan integratsiya qilinishi natijasida til o'zgarishlarini bashorat qilish va lingvistik model yaratish imkonini beradi. Kelgusida ushbu korpusni yanada rivojlantirish, avtomatlashtirish hamda xalqaro miqyosda foydalanish imkoniyatlarini kengaytirish bo'yicha izlanishlar olib borish maqsadga muvofiqdir.

Foydalanilgan adabiyotlar

1. Akademik N. O'zbek tilining leksik rivojlanishi. Toshkent: Akademiya nashriyoti. 2020.
2. Ataboyev N.B. Mediamatnlar diaxron korpusida til rivojining empirik tahlil tamoyillari. Monografiya. Buxoro. 2024.
3. Elov B.B., Hamroyeva Sh.M., Xusainova Z.Y. NLP (Tabiiy tilga ishlov berish)ning vazifalari va zamonaviy yondashuvlar. 2022
4. Biber D. Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press. 1998.
5. Gries, S. T. Quantitative Corpus Linguistics with R: A Practical Introduction. Routledge. 2009.
6. Husainova G. O'zbek tilining lingvistik me'yorlari. Toshkent: Fan. 2018.
7. Hasanova M. 2021. Kompyuter lingvistikasi va NLP texnologiyalari. Toshkent: Fan. 2021.
8. Jurafsky D. and Martin J. H. Speech and Language Processing. Pearson. 2021.
9. Karimov I. O'zbekistonning o'z yo'li – taraqqiyot yo'li. Toshkent: O'zbekiston. 1997.
10. McEnery T. and Hardie A. Corpus Linguistics: Method, Theory and Practice. Cambridge University Press. 2012.

11. Stubbs M. Words and Phrases: Corpus Studies of Lexical Semantics. Blackwell Publishing. 2001.
12. Sinclair J. Corpus, Concordance, Collocation. Oxford University Press. 1991.
13. Xusainova Z. O‘zbek tili milliy korpusi qidiruv tizimini optimallashtirishda lemmatizatsiyadan foydalanish. 2023.

MUSTAQILLIK DAVRI NASHRLARIGA ASOSLANGAN DIAXRON KORPUS YARATISHNING LINGVISTIK TA'MINOTI

Xusainova Zilola

Filologiya fanlari falsafa doktori (PhD).
Alisher Navoiy nomidagi Toshkent davlat
o'zbek tili va adabiyoti universiteti.
E-mail: xusainovazilola@navoiy-uni.uz
ORCID: 0000-0003-4357-7515,

Yangibayeva Surayyo

ToshDO'TAU, Kompyuter lingvistikasi mutaxassisligi
I bosqich magistranti.
E-mail: surayyoyangibayeva4@gmail.com

Annotatsiya. Ushbu maqolada mustaqillik davri nashrlariga asoslangan diaxron korpus yaratish va uning lingvistik ta'minoti masalalari yoritiladi. Tadqiqotda mustaqillikdan keyingi davrda o'zbek tilida yuz bergan o'zgarishlarni o'rganish uchun diaxron korpusning ahamiyati tahlil qilinadi. Shuningdek, tarixiy matnlarni raqamlashtirish, morfologik va sintaktik belgilash texnologiyalari, leksik o'zgarishlarni aniqlash metodlari hamda lingvistik resurslarni yaratish jarayonlari ko'rib chiqiladi. Tadqiqotning dolzarbligi mustaqillik davri nashrlarining tilshunoslik nuqtayi nazaridan tahlil qilinmaganligi bilan bog'liq.

Kalit so'zlar: diaxronik korpus, lingvistik ta'minot, tarixiy matnlar, morfologik tahlil, lemmatizatsiya, normalizatsiya.

Annotation. This article explores the creation of a diachronic corpus based on publications from the independence period and its linguistic support. The study analyzes the significance of a diachronic corpus in examining linguistic changes in the Uzbek language after independence. Additionally, it discusses the digitization of historical texts, morphological and syntactic annotation technologies, methods for identifying lexical changes, and the process of developing linguistic resources. The relevance of the study lies in the fact that publications from the independence period have not yet been analyzed from a linguistic perspective.

Keywords: diachronic corpus, linguistic support, historical texts, morphological analysis, lemmatization, normalization.

Аннотация. В данной статье рассматриваются вопросы создания диахронного корпуса, основанного на изданиях периода независимости, и его лингвистического обеспечения. В исследовании анализируется значение диахронного корпуса для изучения изменений, произошедших в узбекском языке после обретения независимости. Также рассматриваются процессы оцифровки исторических текстов, технологии морфологической и синтаксической разметки, методы выявления лексических изменений и создание лингвистических ресурсов. Актуальность исследования обусловлена тем, что издания периода независимости ранее не подвергались лингвистическому анализу.

Ключевые слова: диахронный корпус, лингвистическое обеспечение, исторические тексты, морфологический анализ, лемматизация, нормализация.

Kirish

Mustaqillikdan keyin o'zbek tilida sezilarli o'zgarishlar yuz berdi. Jumladan, kirill yozuvidan lotin yozuviga o'tish, yangi atamalar paydo bo'lishi, leksik va grammatik tuzilmalarning boyib borishi kuzatildi. Ushbu jarayonlarni tizimli o'rganish va tahlil qilish uchun mustaqillik davri nashrlari asosida diaxronik korpus yaratish dolzarb masala hisoblanadi. Diaxron

korpuslar til o'zgarishlarining dinamikasini kuzatish, matnlarning turli davrlardagi leksik va grammatik o'zgarishlarini aniqlash hamda lingvistik modellash jarayonida muhim vosita bo'lib xizmat qiladi [Rissanen M., 2000]. Ushbu tadqiqot diaxron korpus yaratish jarayonida lingvistik ta'minotning asosiy jihatlarini tahlil qilishga qaratilgan.

Asosiy qism

1. Mustaqillik davri nashrlari va diaxron korpus

Mustaqillik davri matbuot va adabiyotlarining lingvistik o'zgarishlarni aks ettirishdagi ahamiyati katta. Ushbu davrda chop etilgan rasmiy hujjatlar, ilmiy maqolalar, ommaviy axborot vositalari materiallari va adabiy asarlar til o'zgarishlarini kuzatish uchun muhim manba hisoblanadi [Biber D., 1993]. Mana shu ahamiyatlarni inobatga olgan holda o'zbek tilining 33 yillik ya'ni, 1991-yildan 2024-yilgacha bo'lgan *mustaqillik davriga oid* nashrlarini (asosan badiiy asarlar) o'zida jamlagan yana bir diaxron korpus yaratish ishlari olib borilmoqda. Bu kabi diaxron korpuslarda leksik-semantik o'zgarishlar, terminologik yangilanishlar va stilistik transformatsiyalar yaqqol namoyon bo'ladi. Masalan, mustaqillikdan keyin siyosiy va iqtisodiy sohalarida jadal rivojlanish natijasida yangi atamalar va iboralar paydo bo'ldi. Shuningdek, adabiy matnlarda ham milliy o'ziga xoslikni aks ettirish tendensiyalari kuzatilmoqda. Ushbu o'zgarishlarni aniq tahlil qilish uchun diaxron korpusning qo'llanilishi muhim ahamiyat kasb etadi. Bu jarayonda lingvistik modellash, leksik-statistik tahlil va sun'iy intellekt asosida ishlov berish texnologiyalaridan foydalanish diaxron tadqiqotlarni yanada samarali olib borishga yordam beradi. Mustaqillik davriga oid matnlar diaxronik korpusga kiritilsa, o'zbek tilining rivojlanish dinamikasini yanada aniqroq tahlil qilish imkoniyati kengayadi.

2. Lingvistik ta'minot tarkibiy qismlari

Har qanday korpusni yaratishda lingvistik jarayonlar muhim ahamiyat kasb etadi. Bu jarayonlar til o'zgarishlarini tizimli ravishda o'rganish, matnlarni sintaktik va semantik jihatdan chuqur tahlil qilish hamda lingvistik ma'lumotlarni aniq va ishonchli tarzda qayta ishlash imkonini beradi. Bundan tashqari, diaxron korpuslarga ham lingvistik tadqiqotlar uchun mustahkam ma'lumotlar bazasini shakllantirishda asosiy vositalardan biri sifatida qaraladi. Diaxron korpus yaratishda ham quyidagi lingvistik jarayonlar muhim ahamiyatga ega:

2.1. Matnlarni tanlash va tasniflash

Matnlarning stilistik va funksional turlari bo'yicha tasniflanishi, nashr yili va mualliflik omillari hisobga olinishi kerak. Shuningdek, matnlar yozuv tizimi (kirill yoki lotin) bo'yicha ajratilishi lozim. Masalan, Helsinkidagi diaxron korpus tadqiqotlarida matnlarni kategoriyalarga ajratishning samarali usullari qo'llanilgan [Rissanen M., 2000].

2.2. Normalizatsiya va transliteratsiya

Lotin va kirill yozuvidagi matnlarni bir-biriga moslashtirish, eski imlo qoidalarini yangi standartlarga mos ravishda tahlil qilish muhimdir [Jurafsky D. and Martin J.H., 2021]. Masalan, "*xalq xo'jaligi*" iborasi keyingi yillarda "*iqtisodiyot*" so'zi bilan almashtirilgan. Yangi imlo qoidalari bo'yicha eski imlo qoidalarida "*kechayu-kunduz*", "*ertalabki*" shaklida qo'llanilgan so'zlar keying yillarda yangi imlo qoidalariga ko'ra "*kecha-yu kunduz*", "*ertalabgi*" shaklida qo'llanilishi kabilar.

2.3. Morfologik va sintaktik belgilash

Morfologik analizatorlar va avtomatlashtirilgan tahlil usullari orqali so'z shakllari va ularning grammatik xususiyatlari belgilanishi zarur. TreeTagger va UDPipe kabi vositalardan foydalanish mumkin [Kuznetsov I., 2014].

2.4. Leksik o'zgarishlarni kuzatish

Mustaqillikdan keyingi davrda paydo bo'lgan yoki faol qo'llanila boshlagan so'z va iboralar o'rganiladi. Masalan, "*mustaqil davlat*", "*milliy g'oya*", "*ijtimoiy tarmoq*" kabi terminlarning tarixi tahlil qilinadi.

Mana shu ahamiyatlarni o'zida jamlagan holda mustaqillik davri nashrlariga asoslangan diaxron korpusning lingvistik ta'minoti bosqichma-bosqich yaratilmoqda. Lingvistik taminotni yaratishda dastlab 33 yil oralig'ida nashr qilingan badiiy asarlar to'plandi. Ushbu davrda nashr

qilingan matnlar ichidan biz aynan badiiy asarlarni tanlab olishimizdan maqsad badiiy asarlardagi til o'zgarishlarini o'rganish, badiiy asar tilida eskirgan yoki yangidan paydo bo'lgan so'zlarni aniqlash, ya'ni umuman olganda badiiy asar leksikasini o'rganishdir. Shu sababli to'plangan matnlar dastlab janrlarga, nashr qilingan yillarga ko'ra guruhlariga ajratildi. Hozirda to'plangan matnlarni korpus talablariga mos shaklda tozalash ishlari olib borilmoqda. Diaxron korpus yaratishda lingvistik ta'minot quyidagi qamrovlarni o'z ichiga olganini ko'rishimiz mumkin:

1-jadval. Mustaqillik davri nashrlari diaxron korpusi qamrovi

Davr	Yillar oralig'i	Janri	Til doirasi
Mustaqillik davri nashrlari	1991-2024	Badiiy asarlar	Bir tilli (o'zbek)

Ushbu qamrovga asoslangan holda quyidagi matnlar to'plandi (2-jadval). Matnlarni tozalash jarayonida bu raqamlar hali o'zgarish ehtimoli mavjud. Chunki izlanishlar hali davom etmoqda.

2-jadval. To'plangan matnlar

Mualliflar	Badiiy asar	Yillar kesimi	Janri	Formati	Yozuv shakli
110 ta	220 ta	1991-2024	Roman, qissa, hikoya, esse, drama	pdf	Krill, latin,

3. Duch kelinadigan muammolar:

Diaxron korpus yaratishda bir qator texnik, lingvistik va metodologik muammolarga duch kelinadi. Ushbu qiyinchiliklarni hal etish korpusning aniqligi va samaradorligini oshirishga xizmat qiladi.

1-jadval. Diaxron korpus yaratishda yuzaga keladigan muammolar

Matnlarning aksariyat qismi raqamlashtirilmagan:	Mustaqillik davrining dastlabki yillariga oid gazeta va jurnal maqolalarining aksariyati faqat bosma nashr shaklida mavjud.
Yozuv tizimining o'zgarishi:	Lotin-kirill o'zgarishlari lingvistik tahlil jarayonlarini murakkablashtiradi.
Leksik va grammatik o'zgarishlar:	Yangi atamalarining paydo bo'lishi va ba'zi eski iboralarning eskirishi diaxronik korpus tahlilini murakkablashtiradi.

Ushbu biz keltirib o'tgan muammolar diaxron korpus yaratishda bir qancha murakkabliklarni yaratishi mumkin. Xususan, mustaqillik davrining dastlabki yillariga oid gazeta va jurnal maqolalarining aksariyati faqat bosma nashr shaklida mavjudligi raqamlashtirish jarayonining murakkabligini oshiradi va matnlarni OCR (*Optical Character Recognition*) texnologiyalari orqali aniq qayta ishlash muammosini keltirib chiqaradi. Ayniqsa, eski matnlarda ishlatilgan yozuv shriftlari va bosma sifati OCR tizimlarining aniqligiga ta'sir qilishi mumkin. Bu muammoni bartaraf etish uchun yanada ilg'or sun'iy intellekt asosidagi matn tanib olish texnologiyalarini joriy qilish zarur. Shuningdek, raqamlashtirilgan matnlarni verifikatsiya qilish va tahrirlash jarayonlarini avtomatlashtirish uchun lingvistik modellar va neyron tarmoqlar asosida ishlaydigan vositalardan foydalanish maqsadga muvofiqdir. Yozuv tizimining o'zgarishi ham ayniqsa, ma'lumotlar bazasini shakllantirishda bir qancha imloviy va texnik muammolarni keltirib chiqaradi. Bu muammoni bartaraf etishda esa avtomatik transliteratsiya tizimidan foydalanishimiz lozim. Natijalarning aniq bo'lishi uchun esa ushbu tizimlarni yanada rivojlantirish talab qilinadi.

Bundan tashqari korpuslarni yaratishda tabiiy til masalalari bilan ishlashda mashinali o'rganish kabi ham to'siqqa duch kelindi – uning algoritmlari odatda raqamlar ustida amallarni bajaradi. Tabiiy tilda esa matnlar qayta ishlanadi. Shunday qilib, tabiiy tildagi matnni raqamlarga aylantirishga to'g'ri keladi. Ushbu jarayon NLPda *matn vektorizatsiyasi* deb nomlanadi. Ushbu bosqich matnli ma'lumotlarni tahlil qilish uchun mashinali o'rganish jarayonidagi asosiy qadamdir va turli vektorlashtirish algoritmlari yakuniy natijalarga keskin ta'sir qiladi [Elov B.B., Xusainova Z.Y., Xudayberganov N.U. 2023].

Xulosa

Mustaqillik davri nashrlari asosida diaxron korpus yaratish o'zbek tilidagi o'zgarishlarni tizimli o'rganish uchun muhim vositadir. Ushbu tadqiqotda diaxron korpus yaratish jarayonida lingvistik ta'minotning asosiy jihatlari ko'rib chiqildi va mavjud muammolar aniqlandi. Korpus lingvistikasi tamoyillariga asoslangan holda tarixiy matnlarni jamlash, ularni lingvistik jihatdan belgilash va analiz qilish orqali mustaqillik yillarida yuz bergan o'zgarishlarni ilmiy asosda o'rganish imkoniyati yaratildi. Bundan tashqari, diaxron korpusning qo'llanilishi o'zbek tilining rivojlanish tendensiyalarini aniqlashda muhim rol o'ynaydi. Korpus lingvistikasi yondashuvlari asosida leksik va grammatik o'zgarishlarni tizimli tahlil qilish, matnlarni avtomatik tarzda tasniflash va semantik jihatdan o'rganish imkoniyati paydo bo'ladi. Bu esa nafaqat tilshunoslik, balki sun'iy intellekt asosidagi tabiiy tilni qayta ishlash (NLP) sohasida ham katta ahamiyat kasb etadi.

Adabiyotlar:

1. Biber D. Representativeness in corpus design. Literary and Linguistic Computing, 1993. 8(4), 243-257.
2. Elov B.B., Xusainova Z.Y., Xudayberganov N.U. O'zbek tili korpusi matnlari uchun tf-idf statistik ko'rsatkichni hisoblash. Maqola: 2023: 1783-1784
3. Gries S.T. Quantitative Corpus Linguistics with R: A Practical Introduction. Routledge. 2009.
4. Jurafsky D. and Martin J.H. Speech and Language Processing (3rd ed.) Pearson. 2021.
5. Kilgariff A. and Grefenstette G. Introduction to the Special Issue on the Web as Corpus. Computational Linguistics, 2003. 333-347.
6. Kuznetsov I. Russian National Corpus: Structure and Applications. Computational Linguistics and Intelligent Text Processing, 2014. 781-795.
7. McEnery T. and Hardie, A. (2012). Corpus Linguistics: Method, Theory and Practice. Cambridge University Press.
8. Rissanen M. The Helsinki Corpus of English Texts: A Diachronic Perspective on English Syntax. Language and Computers, 2000. 30(1), 21-42.
9. Sinclair J. Corpus, Concordance, Collocation. Oxford University Press. 1991.
10. Stubbs M. Words and Phrases: Corpus Studies of Lexical Semantics. Blackwell Publishing. 2001.

SAVOL-JAVOB TIZIMLARIDA LEMMALASH ALGORITMINING O'RNI

Sharipov Maqsud Siddiqovich

T.f.n., Kompyuter ilmlari kafedrası dotsenti
Abu-Rayhon Beruniy nomidagi Urganch Davlat Universiteti
E-mail: maqsbek72@gmail.com

Sobirov Og'abek Ollayor o'g'li

Kompyuter ilmlari kafedrası tayanch doktoranti
Abu-Rayhon Beruniy nomidagi Urganch Davlat Universiteti
E-mail: sobirovogabek0409@gmail.com

Annotatsiya: Savol-javob tizimlari tabiiy tilni qayta ishlash sohasining muhim yo'nalishlaridan biri bo'lib, inson so'rovlari asosida tegishli javoblarni avtomatik tarzda taqdim etishni maqsad qiladi. Bunday tizimlarning samaradorligi va aniqligi ko'p jihatdan til birliklarini to'g'ri tahlil qilishga bog'liq. Lemmalash - so'zlarni ularning asosiy shakliga keltirish jarayoni bo'lib, savol-javob tizimlarida matnni normallashtirish va semantik tahlil uchun asosiy bosqich hisoblanadi. Ushbu maqolada lemmalash jarayonining savol-javob tizimlari natijalariga ta'siri, uning aniqlik va ishlash samaradorligini oshirishdagi roli tahlil qilinadi. Shuningdek, turli lemmalash usullari va ularning savol-javob tizimlarida qo'llanilishi o'rganiladi.

Kalit so'zlar: Tabiiy tilni qayta ishlash, lemma, lemmalash, savol-javob tizimlari, so'z shakli, normallashtirish.

Абстракт: Системы вопрос-ответ являются важным направлением в области обработки естественного языка и предназначены для автоматического предоставления соответствующих ответов на основе пользовательских запросов. Эффективность и точность таких систем во многом зависят от правильного анализа языковых единиц. Лемматизация - это процесс приведения слов к их базовой форме, который является ключевым этапом нормализации текста и семантического анализа в системах вопрос-ответ. В данной статье анализируется влияние процесса лемматизации на результаты систем вопрос-ответ, а также его роль в повышении точности и эффективности работы. Кроме того, изучаются различные методы лемматизации и их применение в системах вопрос-ответ.

Ключевые слова: Обработка естественного языка, lemma, лемматизация, системы вопрос-ответ, форма слова, нормализация.

Abstract: Question-answer systems are an important direction in the field of natural language processing, aimed at automatically providing relevant answers based on user queries. The efficiency and accuracy of such systems largely depend on the correct analysis of linguistic units. Lemmatization is the process of bringing words to their base form, serving as a key step in text normalization and semantic analysis in question-answer systems. This paper analyzes the impact of the lemmatization process on the performance of question-answer systems and its role in improving accuracy and operational efficiency. Additionally, various lemmatization methods and their applications in question-answer systems are explored.

Keywords: Natural language processing (NLP), lemma, lemmatization, question-answer systems (Q&A), word form, normalization.

Kirish: Hozirgi zamonaviy axborot texnologiyalari rivojlanishi natijasida insonlar tomonidan berilgan tabiiy til so'rovlari asosida avtomatik javob taqdim eta oladigan savol-javob tizimlariga ehtiyoj ortib bormoqda. Ushbu tizimlar Sun'iy Intellekt (ing. Artificial Intelligence - AI) va Tabiiy Tilni Qayta Ishlash (Natural Language Processing - NLP) sohalarining rivojlanishi bilan yanada mukammallashmoqda. Savol-javob tizimlari foydalanuvchilarga katta hajmdagi

ma'lumotlardan tezkor va aniq javob olish imkonini beradi. Bunday tizimlar ilmiy izlanishlar, tibbiyot, huquqshunoslik, ta'lim, elektron tijorat va boshqa ko'plab sohalarda katta ahamiyatga ega.

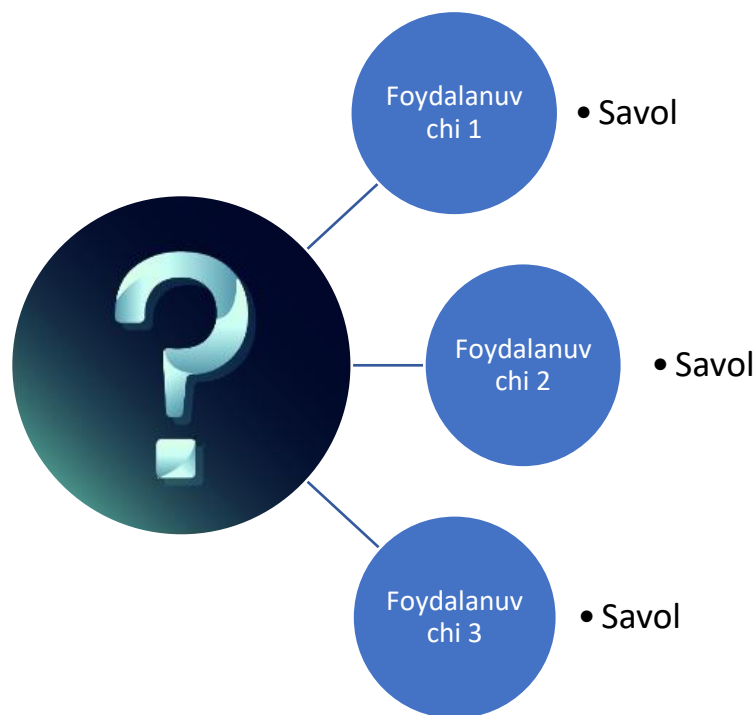
Savol-javob tizimlarini ishlab chiqish jarayonida bir qancha muammolar mavjud bo'lib, ularning asosiylari orasida tabiiy tilning murakkabligi, so'zlarning ko'p ma'noliligi, morfologik o'zgarishlar va kontekstga bog'liq ma'no tafovutlari kiradi. Ayniqsa, turkiy tillar, jumladan, o'zbek tilining morfologik jihatdan boyligi ushbu tizimlar uchun qo'shimcha qiyinchiliklarni keltirib chiqaradi. Savollarning to'g'ri tushunilishi va tegishli javoblarning aniqligi tizim samaradorligini belgilovchi asosiy omillardir.

Lemmalash ushbu muammolarni hal etishda muhim vositalardan biri hisoblanadi. Lemmalash jarayoni so'zlarni ularning asosiy (leksik) shakliga keltirib, tizim tomonidan matnning yaxlit tushunilishini ta'minlaydi. Bu savollarning tahlil qilinishi va tegishli javoblarni aniqlash jarayonini yengillashtiradi hamda tizim natijalarining aniqligini oshiradi. Lemmalash usullarining to'g'ri tanlanishi va qo'llanilishi savol-javob tizimlarining samarali ishlashi uchun muhim hisoblanadi.

Lemmalash — tabiiy tilni qayta ishlashda so'zlarning lug'aviy (asosiy) shaklini aniqlash jarayonidir. Bu jarayon morfologik tahlil yordamida so'zning grammatik qo'shimchalardan tozalangan asosiy ko'rinishini aniqlashga asoslanadi. Lemmalash natijasida bir xil leksik ma'noga ega bo'lgan turli grammatik shakllar bitta asosiy shaklga keltiriladi.

Masalan, quyidagi o'zbekcha so'zlar uchun lemmalash natijalari:

- *kitoblar, kitobning, kitobga* → *kitob*
- *yuritildi, yuritayapti, yuritmoq* → *yuritmoq*
- *o'quvchilar, o'quvchiga, o'quvchini* → *o'quvchi*



Lemmalash ayniqsa, savol-javob tizimlari uchun muhim bo'lib, u turli shakllarda ifodalangan so'zlarni yagona lug'aviy birlik sifatida qayta ishlash imkonini beradi. Bu esa tizimning so'z ma'nolarini to'g'ri tushunishi va mos keluvchi javoblarni aniqlashida muhim ahamiyat kasb etadi.

Ushbu yo'nalishda ko'plab ilmiy tadqiqotlar olib borilgan bo'lib, ular orasida o'zbek tili uchun lemmalash usullarining samaradorligi o'rganilgan (Mahmudjonova, 2022) va qo'shimchalar turlari, ularni ajratib olish (Sobirov, 2024) haqida fikrlar bayon etilgan, bundan

tashqari axborot tizimlarida ma'lumot olishda lemmalashning muhimligi (Elov et al., 2023) ta'kidlangan. Bundan tashqari muallif (Sattarova, 2024) o'z maqolasida o'zbek tilida tabiiy tilni qayta ishlash sohasida uchraydigan muayyan muammolar, xususan, lemmatizatsiya, stemming, tovushni tanib olish va semantik tahlil masalalarini yoritgan. Savol-javob tizimlarining yana bir jihati shundaki, savol berilgan tildan qat'iy nazar javobni istalgan tilde bera olishi kerak, yoki aksincha turli xil tillarda berilgan savolni tushuna olishi zarur. Bunday holatlarda mashina tarjimasiga ehtiyoj tug'iladi, ushbu mualliflar (Sharipov et al., 2024) o'z ishlarida tarjimada lemmalashning muhimligi haqida fikrlar bayon etishgan.

Savol-javob tizimlari borasida olib borilgan ishlarga qaraydigan bo'lsak, mualliflar (Karimberdiyevich, 2024) o'z maqolasida ekspert tizimlarining tuzilishi, jumladan, bilimlar bazasi, xulosa yuritish mexanizmi, foydalanuvchi interfeysi hamda bilim olish va yangilash tizimi haqida ma'lumot berishgan. Ushbu tadqiqot (Musayev et al., 2025) TF-IDF vektor fazo modeli va jumlar o'xshashligiga asoslangan yopiq domenli savol-javob tizimini ishlab chiqishga bag'ishlangan bo'lib, 2000 ta savol-javob juftligidan iborat korpus yig'ilgan, ishlov berilgan va modelni o'qitishda qo'llanilgan bo'lib, matnlar TF-IDF yordamida vektorlashtirilib, kosinus o'xshashligi orqali semantik solishtirish amalga oshirilgan, natijalarga ko'ra, TF-IDF modeli bilan aniqlik 95.2%ga, jumlar o'xshashligi ko'rsatkichlari bilan esa 94.8%ga yetgani aniqlangan. Ushbu maqolada (Yurekli et al., 2008) Anadolu universitetining masofaviy ta'lim oluvchilari uchun turk tilidagi savol-javob tizimini yaratish borasidagi ishlar yoritilgan bo'lib, 205 ta PDF formatdagi elektron kitob matnga aylantirilgan va duch kelgan muammolar va ularning yechimlari bayon qilingan. Tadqiqot (Sarkar et al., 2015) so'rov maqsadini anglaydigan savol-javob tizimini ishlab chiqishga qaratilib, sug'urta sohasi bo'yicha yaratilgan tizim 200 ta so'rovda sinovdan o'tkazilib, ekspertlar tomonidan tasdiqlangan.

Savol-javob tizimi tabiiy tilni qayta ishlash va sun'iy intellekt asosida foydalanuvchilarning savollariga aniq va mantiqiy javob beruvchi dasturiy tizimdir. Umuman olganda, savol-javob tizimi quyidagi asosiy tarkibiy qismlardan iborat:

- Savolni tahlil qilish moduli – foydalanuvchi tomonidan kiritilgan savolning tuzilishi va maqsadini aniqlaydi.
- Ma'lumotlar bazasi yoki bilim bazasi – tizim javob topish uchun foydalanadigan manbalar (matnlar, hujjatlar, grafiklari, veb-resurslar va h.k.).
- Lug'aviy va semantik tahlil – so'zlarning ma'nosini va ularning o'zaro bog'liqligini aniqlash.
- Javobni topish va ishlab chiqish – tizim savolga mos keluvchi javobni tanlaydi yoki generatsiya qiladi.
- Javobni taqdim etish – foydalanuvchiga aniq va tushunarli shaklda javobni yetkazadi.

ChatGPT kabi zamonaviy tizimlar katta hajmdagi matnli ma'lumotlar asosida o'qitilgan neyron tarmoqlar yordamida ishlaydi. Ushbu tizimlarning asosiy xususiyatlari kontekstni tushunish va davom ettirish qobiliyati, murakkab savollarni qayta ishlash va mos javob generatsiya qilish, matnlarning semantik va sintaktik tahlilini amalga oshirish hamda oldindan o'rgatilgan model asosida ma'lumotlarni qayta ishlash va yangi bilimlarni generatsiya qilish imkoniyatlaridan iborat.

Savol-javob tizimlari turli sohalar uchun maxsus ishlab chiqilishi mumkin. Har bir soha uchun tizim o'ziga xos talablarni qondirishi kerak:

- Tibbiyot – kasalliklar, simptomlar va davolash usullari bo'yicha savollarga javob beruvchi tizimlar.
- Huquq – qonunchilik va huquqiy hujjatlar asosida maslahat beruvchi tizimlar.
- Ta'lim – talabalar va o'quvchilar uchun avtomatlashtirilgan maslahat beruvchi tizimlar.
- Elektron tijorat – mahsulotlar, narxlar va mijozlarga xizmat ko'rsatish bo'yicha interaktiv botlar.

Xulosa. Savol-javob tizimlari axborot olish jarayonini tez va samarali amalga oshirish uchun juda muhimdir. Lemmalash va boshqa NLP texnikalari ushbu tizimlarning samaradorligini oshirishda

muhim rol o'ynaydi. Sun'iy intellekt asosidagi tizimlarning rivojlanishi bilan savol-javob tizimlari yanada aniq va foydali bo'lib bormoqda. Ushbu maqolada savol-javob tizimlarining ishlash tamoyillari va NLP usullari bilan bog'liq texnologiyalar tahlil qilindi.

Foydalanilgan adabiyotlar:

1. Elov, B., Khamroeva, S., & Khusainova, Z. (2023). O'ZBEK TILIDAGI TURLI TUZILISHLI SO'ZLARNI LEMMALASH USULLARI. 13–19.
2. Karimberdiyevich, M. O. (2024). EKSPERT TIZIMLARI YARATISH VA ULARNING MUAMMOLARI. ILM-FAN YANGILIKLARI KONFERENSIYASI, 2(1), 123–126.
3. Mahmudjonova, G. (2022). O 'ZBEK TILIDA LEMMATIZATSIYA HODISASI. LEMMATIZATSIYA VA STEMMINGNI FARQLASH. Prospects of Uzbek Applied Philology, 1(1).
4. Musayev, M., Ochilov, M., Kholmatov, O., & Narzullayev, O. (2025). VEKTOR FAZO MODELI HAMDA JUMLALAR O 'XSHASHLIGI O 'LCHOVLARIGA ASOSLANGAN SAVOL-JAVOB TIZIMI ISHLAB CHIQISH. DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE, 3(1), 23–30.
5. Sarkar, S., Rao, V., Mithra, S. M. B., & Rao, S. V. R. K. (2015). NLP algorithm based question and answering system. Seventh International Conference on Computational Intelligence, Modeling and Simulation, 5.
6. Sattarova, S. (2024). Advancing natural language processing in uzbek: challenges and solutions. MODERN PROBLEMS AND PROSPECTS OF APPLIED MATHEMATICS, 1(01).
7. Sharipov, M., Sobirov, O., & Sharipova, G. (2024). Mashinali tarjima tizimlari uchun o'zbek tili so'zlarini lemmalash va gap strukturalarini tahlil qilish. «CONTEMPORARY TECHNOLOGIES OF COMPUTATIONAL LINGUISTICS», 2(22.04), 379–383. Retrieved from <https://www.myscience.uz/index.php/linguistics/article/view/83>
8. Sobirov, O. (2024). O'zbek tilidagi so'zlarning lemmasini topish. COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS, 1(1), 67–71.
9. Yurekli, B., Arslan, A., Senel, H. G., & Yilmazel, O. (2008). Turkish question answering question answering for distance education students. ICSoft 2008 - Proceedings of the 3rd International Conference on Software and Data Technologies, ISDM(ABF/-). doi: 10.5220/0001900703200323

COGNITIVE FEATURES OF ADVERTISING TEXTS WITH VERBAL AND NONVERBAL COMPONENTS

Odilova Kamola Takhirovna

Doctoral student

Uzbekistan State World Languages University

E-mail: kt.odilova@uzswlu.uz

ORCID: <https://orcid.org/0009-0002-3512-1032>

Annotation: This research explores the cognitive features of advertising texts that incorporate both verbal and nonverbal components. The study examines how these multimodal elements interact to create meaning, influence consumer perception, and enhance message retention. By analyzing various advertising campaigns across different media, the paper highlights the importance of integrating linguistic and visual elements in effective advertising communication. The findings contribute to our understanding of cognitive processing in multimodal advertising and offer insights for marketers and advertisers in crafting more impactful messages.

Keywords: advertising linguistics, multimodal discourse, cognitive linguistics, visual semiotics, consumer psychology, brand communication, persuasive techniques

Annotatsiya: Ushbu tadqiqot verbal va noverbal komponentlarni o'z ichiga olgan reklama matnlarining kognitiv xususiyatlarini o'rganadi. Ishda ushbu multimodal elementlarning ma'no yaratish, iste'molchilar idrokiga ta'sir ko'rsatish va xabarlarning esda qolib ketishini yaxshilash uchun qanday o'zaro ta'sir qilishi tahlil qilinadi. Turli ommaviy axborot vositalarida turli reklama kampaniyalarini tahlil qilish asosida maqola samarali reklama kommunikatsiyasida lingvistik va vizual elementlarni birlashtirish muhimligini ta'kidlaydi. Tadqiqot natijalari multimodal reklamada kognitiv qayta ishlashni tushunishga yordam beradi va marketologlar va reklama beruvchilarga yanada samarali xabarlarni yaratish uchun g'oyalarni taklif qiladi.

Kalit so'zlar: reklama lingvistikasi, multimodal diskurs, kognitiv lingvistika, vizual semiotika, iste'molchi psixologiyasi, brend-kommunikatsiya, ishontirish texnikasi

Аннотация: Данное исследование рассматривает когнитивные особенности рекламных текстов, включающих как вербальные, так и невербальные компоненты. В работе анализируется, как эти мультимодальные элементы взаимодействуют для создания смысла, влияния на восприятие потребителей и улучшения запоминаемости сообщений. На основе анализа различных рекламных кампаний в разных медиа, статья подчеркивает важность интеграции лингвистических и визуальных элементов в эффективной рекламной коммуникации. Результаты исследования способствуют пониманию когнитивной обработки в мультимодальной рекламе и предлагают маркетологам и рекламодателям идеи для создания более эффективных сообщений.

Ключевые слова: рекламная лингвистика, мультимодальный дискурс, когнитивная лингвистика, визуальная семиотика, психология потребителя, бренд-коммуникация, техники убеждения

Advertising has become an integral part of modern society, permeating various aspects of our daily lives. As a complex form of communication, advertising relies heavily on both verbal and nonverbal components to convey messages, evoke emotions, and influence consumer behavior. This study aims to explore the cognitive features of advertising texts that incorporate these multimodal elements, focusing on how they interact to create meaning and impact the audience's perception and decision-making processes.

The cognitive approach to advertising analysis recognizes that human information processing involves multiple sensory inputs and mental representations. In the context of advertising, this means that consumers do not simply read or view ads; they engage in a complex cognitive process that involves perception, attention, memory, and decision-making. By understanding these cognitive mechanisms, advertisers can create more effective and persuasive messages that resonate with their target audience.

One of the key aspects of cognitive processing in advertising is the integration of verbal and nonverbal components. Verbal elements, such as headlines, body copy, and slogans, provide explicit information and arguments. Nonverbal components, including images, colors, typography, and layout, contribute to the overall message through implicit associations and emotional appeal. The interplay between these elements creates a holistic advertising experience that can be more powerful than either component alone [Forceville C., 2020: 78].

Research has shown that the human brain processes visual information much faster than text, making nonverbal elements crucial in capturing attention and conveying messages quickly. For instance, a study by MIT neuroscientists found that the brain can process images seen for as little as 13 milliseconds. This rapid processing of visual information underscores the importance of carefully crafted visual elements in advertising to make an immediate impact on the viewer [Potter M. C., et al., 2014: 270].

However, the verbal component remains essential in providing context, elaboration, and persuasive arguments. The cognitive linguistic theory of conceptual metaphor, developed by Lakoff and Johnson, offers valuable insights into how language in advertising can shape consumer perceptions. According to this theory, metaphors are not merely linguistic devices but fundamental cognitive structures that influence how we understand and experience the world. Advertisers often exploit conceptual metaphors to create associations between products and desirable concepts or experiences [Lakoff G. & Johnson M., 1980: 124].

For example, the conceptual metaphor "LIFE IS A JOURNEY" is frequently used in advertising to position products as facilitators or companions on this metaphorical journey. Car advertisements often employ this metaphor, presenting vehicles not just as modes of transportation but as enablers of life experiences and personal growth. The verbal component, in conjunction with carefully selected imagery, reinforces this conceptual mapping and enhances the overall persuasive effect of the advertisement.

The integration of verbal and nonverbal elements in advertising also plays a crucial role in brand identity and recognition. Consistent use of visual elements such as logos, color schemes, and typographic styles, combined with distinctive verbal elements like slogans or taglines, creates a cohesive brand image that is easily recognizable and memorable. This multimodal approach to branding leverages the cognitive processes of pattern recognition and associative memory, allowing consumers to quickly identify and recall brand-related information [Kress G. & van Leeuwen T., 2006: 201].

Another important cognitive feature of advertising texts with verbal and nonverbal components is their ability to create emotional responses. Emotions play a significant role in decision-making and memory formation, making them a powerful tool in advertising. While verbal elements can describe emotions or evoke them through storytelling, nonverbal components like images and music can trigger immediate emotional responses. The combination of these elements can create a more immersive and emotionally resonant advertising experience [Damasio A. R., 1994: 167].

For instance, a public service announcement about environmental conservation might pair stark images of polluted landscapes with compelling verbal messages about the urgency of action. The visual elements evoke immediate emotional responses such as concern or sadness, while the verbal component provides context and calls to action. This multimodal approach engages both the emotional and rational aspects of cognition, potentially leading to stronger persuasive effects and behavioral changes.

The cognitive processing of multimodal advertising texts also involves the concept of intertextuality, where meaning is derived not just from the ad itself but from its relationships with other texts and cultural references. Advertisers often leverage this cognitive feature by incorporating familiar visual or verbal elements that resonate with the target audience's shared cultural knowledge. This intertextual approach can create a sense of familiarity and engagement, as consumers enjoy recognizing and decoding these references [Kristeva J., 1980: 69].

For example, a print ad might visually reference a famous painting while using a headline that plays on a well-known literary quote. This combination of visual and verbal intertextuality engages the viewer's cognitive processes of recognition and interpretation, potentially increasing the ad's memorability and effectiveness [Cook G., 2001: 189].

The cognitive features of advertising texts with verbal and nonverbal components are also influenced by the specific medium through which they are delivered. Different media formats – such as print, television, digital, or outdoor advertising – offer unique opportunities and constraints for integrating verbal and nonverbal elements. These medium-specific characteristics affect how consumers cognitively process the advertising message [McLuhan M., 1964: 7].

For instance, digital advertising allows for interactive and dynamic integration of verbal and nonverbal elements, enabling advertisers to create more engaging and personalized experiences. A banner ad might combine animated visuals with changing text to tell a story or demonstrate a product's features. This dynamic presentation can capture and maintain attention, potentially leading to better information processing and retention [Sundar S. S. & Kim J., 2005: 585].

The cognitive load theory, developed by John Sweller, provides another important perspective on how multimodal advertising texts are processed. According to this theory, our working memory has limited capacity, and cognitive overload can occur when too much information is presented simultaneously. In the context of advertising, this suggests that the integration of verbal and nonverbal elements should be carefully balanced to optimize cognitive processing and avoid overwhelming the viewer [Sweller J., 1988: 257].

For example, the conceptual metaphor "LIFE IS A JOURNEY" is frequently used in advertising to position products as facilitators or companions on this metaphorical journey. Car advertisements often employ this metaphor, presenting vehicles not just as modes of transportation but as enablers of life experiences and personal growth. The verbal component, in conjunction with carefully selected imagery, reinforces this conceptual mapping and enhances the overall persuasive effect of the advertisement.

The integration of verbal and nonverbal elements in advertising also plays a crucial role in brand identity and recognition. Consistent use of visual elements such as logos, color schemes, and typographic styles, combined with distinctive verbal elements like slogans or taglines, creates a cohesive brand image that is easily recognizable and memorable. This multimodal approach to branding leverages the cognitive processes of pattern recognition and associative memory, allowing consumers to quickly identify and recall brand-related information [Kress G. & van Leeuwen T., 2006: 201].

Another important cognitive feature of advertising texts with verbal and nonverbal components is their ability to create emotional responses. Emotions play a significant role in decision-making and memory formation, making them a powerful tool in advertising. While verbal elements can describe emotions or evoke them through storytelling, nonverbal components like images and music can trigger immediate emotional responses. The combination of these elements can create a more immersive and emotionally resonant advertising experience [Damasio A. R., 1994: 167].

For instance, a public service announcement about environmental conservation might pair stark images of polluted landscapes with compelling verbal messages about the urgency of action. The visual elements evoke immediate emotional responses such as concern or sadness, while the verbal component provides context and calls to action. This multimodal approach engages both

the emotional and rational aspects of cognition, potentially leading to stronger persuasive effects and behavioral changes.

The cognitive processing of multimodal advertising texts also involves the concept of intertextuality, where meaning is derived not just from the ad itself but from its relationships with other texts and cultural references. Advertisers often leverage this cognitive feature by incorporating familiar visual or verbal elements that resonate with the target audience's shared cultural knowledge. This intertextual approach can create a sense of familiarity and engagement, as consumers enjoy recognizing and decoding these references [Kristeva J., 1980: 69].

For example, a print ad might visually reference a famous painting while using a headline that plays on a well-known literary quote. This combination of visual and verbal intertextuality engages the viewer's cognitive processes of recognition and interpretation, potentially increasing the ad's memorability and effectiveness [Cook G., 2001: 189].

The cognitive features of advertising texts with verbal and nonverbal components are also influenced by the specific medium through which they are delivered. Different media formats – such as print, television, digital, or outdoor advertising – offer unique opportunities and constraints for integrating verbal and nonverbal elements. These medium-specific characteristics affect how consumers cognitively process the advertising message [McLuhan M., 1964: 7].

For instance, digital advertising allows for interactive and dynamic integration of verbal and nonverbal elements, enabling advertisers to create more engaging and personalized experiences. A banner ad might combine animated visuals with changing text to tell a story or demonstrate a product's features. This dynamic presentation can capture and maintain attention, potentially leading to better information processing and retention [Sundar S. S. & Kim J., 2005: 585].

The cognitive load theory, developed by John Sweller, provides another important perspective on how multimodal advertising texts are processed. According to this theory, our working memory has limited capacity, and cognitive overload can occur when too much information is presented simultaneously. In the context of advertising, this suggests that the integration of verbal and nonverbal elements should be carefully balanced to optimize cognitive processing and avoid overwhelming the viewer [Sweller J., 1988: 257].

Effective advertisements often strike a balance between providing sufficient information and maintaining simplicity. For example, a minimalist print ad might feature a striking image with a brief, impactful headline. This approach reduces cognitive load while still conveying a powerful message. On the other hand, more complex products or services might require a greater amount of verbal information, necessitating a careful design that guides the viewer's attention and facilitates information processing.

The concept of multimodal cohesion is another significant cognitive feature of advertising texts with verbal and nonverbal components. This refers to how different semiotic modes (such as text, image, color, and layout) work together to create a unified message. Cohesion between verbal and nonverbal elements enhances the overall comprehension and impact of the advertisement by reducing cognitive effort and increasing the clarity of the message [van Leeuwen T., 2005: 179].

For instance, in a car advertisement, the verbal description of the vehicle's advanced safety features might be visually reinforced by images showing the car's robust structure or its performance in crash tests. This multimodal cohesion helps to create a more compelling and believable narrative about the car's safety, as the verbal and visual elements mutually support and enhance each other [Bateman J. A., 2014: 243].

The cognitive process of attention allocation is also crucial in understanding how consumers interact with multimodal advertising texts. In a media-saturated environment, capturing and maintaining attention is a significant challenge for advertisers. The integration of verbal and nonverbal elements can be strategically used to guide attention and facilitate information processing [Kahneman D., 1973: 98].

Visual elements, such as bright colors, unusual shapes, or human faces, can serve as attention-grabbing devices, drawing the viewer's eye to specific parts of the advertisement. Once attention is captured, verbal elements can provide more detailed information or persuasive arguments. This interplay between attention-grabbing visuals and informative text leverages the cognitive processes of selective attention and information integration [Pieters R. & Wedel M., 2004: 36].

The concept of cognitive fluency also plays a role in how multimodal advertising texts are processed. Cognitive fluency refers to the ease with which information is processed by the brain. Advertisements that are easy to process are often perceived more positively and are more likely to be remembered. The integration of verbal and nonverbal elements can enhance cognitive fluency by presenting information in multiple, complementary formats [Alter A. L. & Oppenheimer D. M., 2009: 219].

For example, an infographic in an advertisement combines visual representations of data with concise textual explanations. This multimodal presentation can make complex information more accessible and easier to process, potentially leading to better comprehension and retention of the advertising message.

The cognitive feature of embodied cognition is also relevant to understanding how consumers interact with multimodal advertising texts. This theory suggests that our cognitive processes are deeply influenced by our physical experiences and interactions with the world. In advertising, this can be leveraged by creating sensory-rich experiences that engage multiple cognitive modalities [Barsalou L. W., 2008: 617].

For instance, a food advertisement might combine mouth-watering visuals with descriptive text that evokes taste and smell. This multimodal approach activates sensory memories and associations, potentially creating a more vivid and persuasive advertising experience. Similarly, advertisements for tactile products like clothing or furniture might use textures in their visual design, complemented by verbal descriptions of comfort or quality, to engage the viewer's sense of touch [Krishna A., 2012: 332].

The cognitive processes involved in narrative comprehension are also relevant to multimodal advertising texts, particularly in the context of storytelling advertisements. These ads often combine verbal narration with visual storytelling elements to create engaging narratives that resonate with viewers. The integration of verbal and nonverbal elements in narrative ads can facilitate the cognitive processes of story comprehension, character identification, and emotional engagement [Mar R. A., 2004: 1414].

For example, a video advertisement might tell a story through a series of visual scenes, supported by voiceover narration or on-screen text. This multimodal narrative approach engages multiple cognitive processes, including visual processing, language comprehension, and narrative construction, potentially leading to stronger emotional connections and better message retention.

The concept of cognitive dissonance, introduced by Leon Festinger, is another important consideration in the design of multimodal advertising texts. Cognitive dissonance occurs when an individual experiences conflicting attitudes, this concept can be strategically used to create tension or surprise that engages the viewer's attention and promotes deeper processing of the message [Festinger L., 1957: 3].

In multimodal advertising texts, cognitive dissonance can be created through the juxtaposition of contrasting verbal and nonverbal elements. For instance, an environmental awareness campaign might pair beautiful nature imagery with stark text describing pollution statistics. This dissonance between the visual and verbal elements can create a powerful emotional impact and motivate the viewer to reconcile the conflicting information, potentially leading to attitude or behavior change.

The cognitive feature of dual coding theory, proposed by Allan Paivio, is particularly relevant to understanding the effectiveness of multimodal advertising texts. This theory suggests that the human mind processes and stores verbal and visual information separately but in

interconnected systems. When information is presented in both verbal and visual formats, it has a better chance of being remembered because it is encoded in two different ways [Paivio A., 1986: 53].

In advertising, this principle can be leveraged by ensuring that key messages are conveyed through both text and images. For example, a car advertisement might show an image of the vehicle driving through a challenging terrain while the text describes its powerful engine and advanced suspension system. This dual presentation of information increases the likelihood of the message being encoded and recalled later.

The concept of cognitive load, introduced earlier, can be further explored in the context of multimodal advertising texts. While the integration of verbal and nonverbal elements can enhance comprehension and retention, it's crucial to balance the amount and complexity of information presented to avoid overwhelming the viewer's cognitive resources [Sweller J., 1988: 257].

Effective multimodal advertisements often employ strategies to manage cognitive load. This might include using visual hierarchies to guide attention, breaking complex information into digestible chunks, or using consistent visual motifs to create a sense of coherence across different elements of the ad. The goal is to create an advertisement that is informative and engaging without being cognitively taxing.

The role of emotional processing in the cognitive features of multimodal advertising texts is also significant. Emotions play a crucial role in decision-making and memory formation, and advertisers often aim to evoke specific emotional responses through their campaigns. The integration of verbal and nonverbal elements can be particularly effective in eliciting emotional responses [Damasio A. R., 1994: 127].

Visual elements in advertisements can quickly trigger emotional responses through color, imagery, and facial expressions. These emotional cues can then be reinforced or contextualized by the verbal elements. For instance, an advertisement for a charity might use images of people in need to evoke empathy, while the accompanying text provides a rational argument for why donation is important. This combination of emotional and rational appeals leverages both the fast, intuitive processing of visual information and the slower, more deliberate processing of verbal information.

The concept of priming, a cognitive process where exposure to one stimulus influences the response to a subsequent stimulus, is another important consideration in multimodal advertising. In the context of advertising, priming can be used to create associations between the product or brand and certain desirable attributes or emotions [Tulving E. & Schacter D. L., 1990: 301].

Multimodal advertisements can use both verbal and nonverbal elements to create priming effects. For example, a luxury car advertisement might use sophisticated color schemes and elegant typography (nonverbal elements) alongside words like "prestige" and "excellence" (verbal elements) to prime associations with high status and quality. These priming effects can influence how consumers perceive and evaluate the advertised product.

The cognitive process of categorization also plays a role in how consumers interpret multimodal advertising texts. Categorization is the mental process of grouping similar objects or concepts together, which helps in organizing and understanding new information. In advertising, both verbal and nonverbal elements can be used to trigger specific categorizations that influence how the product or brand is perceived [Rosch E., 1978: 27].

For instance, an organic food brand might use earthy colors and natural textures in its visual design, combined with words like "pure" and "natural" in its copy. This multimodal approach aims to categorize the brand within the "natural and healthy" category in the consumer's mind, influencing their perception and potentially their purchasing decisions.

The concept of cognitive fluency, mentioned earlier, can be further explored in relation to the aesthetic aspects of multimodal advertising texts. Cognitive fluency refers to the ease with which information is processed, and research has shown that people tend to prefer and trust

information that is easy to process [Reber R. et al., 2004: 364]. In advertising, this principle can be applied to both verbal and nonverbal elements.

Visually, advertisements that are clean, well-organized, and adhere to established design principles are likely to be processed more fluently. This might include using appropriate white space, consistent color schemes, and clear visual hierarchies. Verbally, cognitive fluency can be enhanced through the use of simple language, familiar phrases, and easy-to-read fonts. When verbal and nonverbal elements are designed with cognitive fluency in mind, the overall advertisement becomes more appealing and persuasive.

The role of working memory in processing multimodal advertising texts is another important cognitive feature to consider. Working memory is the cognitive system responsible for temporarily holding and manipulating information. It has limited capacity, which has implications for how much information can be effectively processed from an advertisement at any given time [Baddeley A., 1992: 556].

In multimodal advertising, the integration of verbal and nonverbal elements can help to optimize the use of working memory. Visual elements can quickly convey complex information that might take many words to describe, freeing up verbal working memory resources. Conversely, text can provide specific details or context that might be difficult to convey visually. By strategically combining these elements, advertisers can communicate more information within the constraints of working memory capacity.

The concept of cognitive schemas, organized patterns of thought or behavior that categorize and interpret information, is also relevant to understanding how consumers process multimodal advertising texts. Schemas help individuals to quickly make sense of new information by relating it to existing knowledge structures [Bartlett F. C., 1932: 201].

In advertising, both verbal and nonverbal elements can be used to activate or create schemas that influence how the product or brand is perceived. For example, an advertisement for a new technology product might use futuristic imagery and technical jargon to activate schemas associated with innovation and advanced technology. These schemas then guide the consumer's interpretation of the product, potentially influencing their attitudes and purchase intentions.

The cognitive process of elaboration is another important feature in the context of multimodal advertising texts. Elaboration refers to the extent to which individuals engage in thoughtful consideration of the information presented to them. High levels of elaboration are associated with stronger and more enduring attitudes [Petty R. E. & Cacioppo J. T., 1986: 123].

Multimodal advertisements can encourage elaboration by presenting information in ways that engage multiple cognitive processes. For instance, an advertisement might use an intriguing visual metaphor that requires some interpretation, paired with text that provides additional context or poses thought-provoking questions. This combination of elements encourages the viewer to engage more deeply with the advertisement, potentially leading to stronger and more lasting impressions.

The concept of cognitive dissonance, introduced earlier, can be further explored in terms of its potential positive effects in advertising. While cognitive dissonance is often seen as a negative experience that individuals seek to resolve, it can also be used strategically in advertising to create engagement and memorability [Festinger L., 1957: 31].

In multimodal advertising texts, cognitive dissonance can be created through the strategic mismatch of verbal and nonverbal elements. For example, a public health campaign might pair pleasant, appealing imagery with stark statistics about health risks. This dissonance between the visual and verbal elements can create a sense of surprise or unease that captures attention and encourages deeper processing of the message. When used judiciously, this technique can lead to more impactful and memorable advertisements.

The role of cognitive heuristics in processing multimodal advertising texts is another important consideration. Heuristics are mental shortcuts or rules of thumb that individuals use to make decisions quickly, especially in situations of uncertainty or when faced with complex

information [Tversky A. & Kahneman D., 1974: 1124]. In the context of advertising, both verbal and nonverbal elements can be designed to trigger specific heuristics that influence consumer decision-making.

For example, the availability heuristic, which leads people to judge the probability of an event based on how easily they can recall relevant examples, can be leveraged in multimodal advertising. An advertisement might use striking visual imagery paired with memorable slogans to make certain product benefits more mentally available to consumers. Similarly, the representativeness heuristic, where people judge the probability of something based on how closely it resembles their mental prototype, can be utilized by carefully crafting both visual and verbal elements to align with consumers' expectations of a particular product category.

The concept of perceptual fluency, closely related to cognitive fluency, is particularly relevant to the nonverbal aspects of multimodal advertising texts. Perceptual fluency refers to the ease with which people can identify and recognize physical characteristics of stimuli. In advertising, this can be applied to elements such as logo design, color schemes, and overall visual layout [Reber R. et al., 2004: 365].

Advertisements that are perceptually fluent are processed more quickly and tend to be evaluated more positively. This can be achieved through consistent use of visual branding elements across different advertisements, making the brand more easily recognizable. When combined with cognitively fluent verbal elements, perceptually fluent visual designs can create powerful, easily processed multimodal advertisements.

The role of cognitive load in multimodal advertising texts warrants further discussion. While it's important to manage cognitive load to avoid overwhelming viewers, there can be strategic benefits to introducing a certain level of complexity. The concept of "desirable difficulties" suggests that introducing some level of challenge in information processing can lead to deeper engagement and better long-term retention [Bjork R. A., 1994: 185].

In multimodal advertising, this might involve creating advertisements that require some level of interpretation or problem-solving. For example, an advertisement might use a visual puzzle that is solved by reading the accompanying text, or vice versa. This approach can increase engagement and memorability, provided the level of difficulty is carefully calibrated to the target audience and context.

The concept of cognitive mapping, which refers to the mental representation of spatial environments, can also be applied to understanding how consumers navigate and interpret multimodal advertising texts. Just as people create mental maps of physical spaces, they also create cognitive maps of information spaces, including advertisements [Tolman E. C., 1948: 189].

In multimodal advertising, the spatial arrangement of verbal and nonverbal elements can influence how information is mentally organized and recalled. For example, the use of visual hierarchies can guide attention and create a mental structure for the information presented. Similarly, the strategic placement of text in relation to images can influence how connections are made between different pieces of information. Understanding and leveraging these cognitive mapping processes can help create more effective and memorable advertisements.

The role of cognitive priming in multimodal advertising texts is another important consideration. Priming refers to the activation of certain associations in memory, which can influence subsequent thoughts and behaviors. In advertising, both verbal and nonverbal elements can be used to prime specific concepts or emotions [Bargh J. A. et al., 1996: 230].

For example, an advertisement might use imagery associated with luxury and exclusivity, priming concepts of high status and quality. This visual priming can then be reinforced by carefully chosen words and phrases. The combined effect of visual and verbal priming can create a powerful influence on how consumers perceive and evaluate the advertised product or brand.

The concept of cognitive dissonance reduction strategies is also relevant to multimodal advertising texts. When faced with cognitive dissonance, individuals typically seek to reduce it

through various strategies, such as changing their attitudes, seeking new information, or reinterpreting the dissonant elements [Festinger L., 1957: 31].

In advertising, this process can be leveraged by presenting information in a way that creates mild cognitive dissonance and then offering a resolution. For example, an advertisement might present a common problem (creating dissonance) and then offer a solution through the advertised product (resolving the dissonance). This approach can be particularly effective when verbal and nonverbal elements work in tandem to create and resolve the dissonance.

The concept of embodied cognition, mentioned earlier, deserves further exploration in the context of multimodal advertising texts. This theory posits that our cognitive processes are deeply rooted in our physical experiences and interactions with the world. In advertising, this suggests that messages that engage multiple senses and simulate physical experiences can be particularly effective [Barsalou L. W., 2008: 617].

Multimodal advertisements can leverage embodied cognition by creating sensory-rich experiences that go beyond just visual and verbal elements. For example, print advertisements might use textured paper to simulate the feel of a product, or digital advertisements might incorporate interactive elements that mimic physical interactions with the product. By engaging multiple senses and tapping into embodied cognitive processes, these advertisements can create more vivid and memorable experiences for consumers.

The role of mental simulation in processing multimodal advertising texts is another important cognitive feature to consider. Mental simulation refers to the process of imagining or mentally rehearsing an action or experience. Research has shown that both verbal descriptions and visual imagery can trigger mental simulations, which can influence attitudes and behaviors [Escalas J. E., 2004: 37].

In multimodal advertising, verbal and nonverbal elements can work together to facilitate mental simulation. For example, an advertisement for a vacation destination might combine evocative imagery with descriptive text that encourages viewers to imagine themselves in the scene. This combination can create a more immersive and persuasive experience than either element alone.

The concept of cognitive schemas, or mental frameworks that help organize and interpret information, is also relevant to understanding how consumers process multimodal advertising texts. Schemas influence how new information is perceived, interpreted, and remembered. In advertising, both verbal and nonverbal elements can activate and interact with existing schemas [Bartlett F. C., 1932: 201].

For example, an advertisement might use visual cues that activate a particular schema (e.g., luxury, family, adventure), which then influences how the verbal content is interpreted. Conversely, verbal elements might prime certain schemas that affect how visual elements are perceived. Understanding and leveraging these schema interactions can help create more effective and resonant multimodal advertisements.

The role of cognitive fluency in aesthetic appreciation is another aspect worth exploring in the context of multimodal advertising texts. Research has shown that people tend to prefer stimuli that are easy to process, a phenomenon known as the "beauty-in-averageness" effect [Reber R. et al., 2004: 364]. This preference for fluency can extend to both verbal and nonverbal elements in advertising.

In multimodal advertisements, cognitive fluency can be enhanced by ensuring that verbal and nonverbal elements are congruent and mutually reinforcing. For example, using fonts that match the mood conveyed by the imagery, or ensuring that the language style aligns with the visual aesthetic. This congruence not only makes the advertisement easier to process but can also enhance its perceived attractiveness and effectiveness.

The concept of cognitive load theory, introduced earlier, can be further applied to understanding how consumers process complex multimodal advertisements. While managing cognitive load is important to avoid overwhelming viewers, there's also evidence that a certain

level of complexity can be beneficial in terms of engagement and memorability [Sweller J., 1988: 257].

In multimodal advertising, this might involve creating a balance between simplicity and complexity. For example, an advertisement might use a visually complex image that requires some interpretation, paired with simple, clear text that guides this interpretation. This approach can create a more engaging and memorable experience than an overly simplistic advertisement, provided the level of complexity is appropriate for the target audience and context.

The role of intertextuality in cognitive processing of multimodal advertising texts is another important consideration. Intertextuality refers to the shaping of a text's meaning by another text, and in multimodal advertising, this can involve references to other advertisements, cultural phenomena, or shared knowledge [Kristeva J., 1980: 69].

In multimodal advertising, intertextuality can manifest through both verbal and nonverbal elements. For example, an advertisement might visually reference a famous painting while using a tagline that alludes to a well-known quote. This interplay of intertextual references across modalities can create rich, layered meanings that engage viewers on multiple cognitive levels.

The effectiveness of intertextuality in multimodal advertising depends on the viewer's ability to recognize and interpret these references. When successful, it can create a sense of shared knowledge between the advertiser and the consumer, enhancing engagement and fostering a positive attitude towards the brand. However, advertisers must carefully consider their target audience to ensure that intertextual references are accessible and meaningful.

The concept of multimodal metaphor is another important cognitive feature in advertising texts with verbal and nonverbal components. While conceptual metaphors have long been recognized as powerful tools in verbal communication, multimodal metaphors extend this concept to include visual and other nonverbal elements [Forceville C., 2009: 19].

In multimodal advertising, metaphors can be created through the interaction of verbal and nonverbal elements. For example, an image might visually represent an abstract concept, while the accompanying text provides the context needed to interpret the metaphor. These multimodal metaphors can be particularly effective in communicating complex ideas or emotional concepts in a concise and memorable way.

The cognitive processing of multimodal metaphors involves integrating information from different sensory modalities and mapping this onto abstract concepts. This process can lead to deeper engagement with the advertisement and more elaborate processing of the message, potentially enhancing its persuasive impact.

The role of visual rhetoric in multimodal advertising texts is another area that merits attention. Visual rhetoric refers to the way visual elements are used to communicate, persuade, and create meaning. In multimodal advertising, visual rhetoric often works in conjunction with verbal rhetoric to create a cohesive and persuasive message [Scott L. M., 1994: 252].

Visual rhetorical devices such as metonymy, synecdoche, and irony can be used alongside verbal elements to create complex meanings and evoke specific responses from viewers. For example, an advertisement might use a visual synecdoche (where a part represents the whole) in conjunction with a verbal metaphor to create a multilayered message that engages viewers on multiple cognitive levels.

The cognitive processing of visual rhetoric in multimodal advertising involves interpreting visual symbols, understanding cultural conventions, and integrating this information with verbal content. This process can lead to more elaborate and memorable engagement with the advertisement, potentially enhancing its effectiveness.

The concept of multimodal framing is another important cognitive feature in advertising texts with verbal and nonverbal components. Framing refers to the way information is presented and how this presentation influences interpretation and decision-making [Entman R. M., 1993: 51].

In multimodal advertising, framing can occur through both verbal and nonverbal elements, and the interaction between these elements can create powerful framing effects. For example, the visual composition of an advertisement might frame a product in a certain way (e.g., as luxurious or practical), while the verbal content reinforces or potentially contradicts this framing.

The cognitive processing of multimodal framing involves integrating and reconciling frames presented through different modalities. This can lead to complex interpretations and potentially influence consumer attitudes and behaviors in ways that single-modality framing might not achieve.

The role of cognitive dissonance in multimodal advertising texts, introduced earlier, can be further explored in terms of its potential for creating memorable and persuasive messages. Cognitive dissonance occurs when a person holds contradictory beliefs or experiences a conflict between their beliefs and actions [Festinger L., 1957: 3].

In multimodal advertising, cognitive dissonance can be strategically created through the juxtaposition of conflicting verbal and nonverbal elements. For example, an advertisement might present a visually appealing image of an unhealthy food product alongside text highlighting its negative health effects. This intentional creation of cognitive dissonance can prompt viewers to engage more deeply with the advertisement as they attempt to resolve the conflict.

The cognitive processing of such dissonance-inducing advertisements involves reconciling conflicting information from different modalities. This process can lead to increased attention, elaboration, and potentially behavior change as viewers seek to resolve the dissonance. However, advertisers must carefully balance the use of cognitive dissonance to avoid alienating or confusing their audience.

The concept of multimodal affordances is another important consideration in understanding the cognitive features of advertising texts with verbal and nonverbal components. Affordances refer to the possibilities for action that an environment or object offers to an individual [Gibson J. J., 1979: 127]. In the context of multimodal advertising, different modalities offer different affordances for meaning-making and interaction.

For example, visual elements might afford quick emotional responses and spatial understanding, while verbal elements afford more detailed and abstract information processing. The integration of these different affordances in multimodal advertising can create rich, multidimensional messages that engage viewers on multiple cognitive levels.

The cognitive processing of multimodal affordances involves recognizing and utilizing the possibilities offered by different modalities. This can lead to more flexible and adaptive engagement with advertising messages, potentially enhancing their effectiveness across different contexts and audiences.

The role of cognitive schemas in interpreting multimodal advertising texts is another area worthy of exploration. Schemas are mental frameworks that help individuals organize and interpret information based on prior knowledge and experiences [Bartlett F. C., 1932: 201]. In multimodal advertising, both verbal and nonverbal elements can activate and interact with these schemas.

For example, a visual element might activate a particular schema (e.g., luxury, family, adventure), which then influences how the accompanying verbal content is interpreted. Conversely, verbal content might prime certain schemas that affect the interpretation of visual elements. The interplay between verbal and nonverbal elements in activating and modifying schemas can create complex and nuanced meanings in advertising messages.

The cognitive processing of schema-activating multimodal advertisements involves integrating new information with existing mental frameworks. This process can lead to more efficient processing of advertising messages, but it can also be used strategically to challenge or modify existing schemas, potentially leading to attitude or behavior change.

The concept of multimodal argumentation is another important cognitive feature in advertising texts with verbal and nonverbal components. While traditional argumentation theory

has focused primarily on verbal discourse, multimodal argumentation extends this to include visual and other nonverbal elements in the construction of persuasive messages [Tseronis A. & Forceville C., 2017: 2].

In multimodal advertising, arguments can be constructed through the interaction of verbal and nonverbal elements. For example, a visual element might serve as evidence for a verbally stated claim, or a series of images might construct a visual argument that is reinforced or elaborated by accompanying text. These multimodal arguments can be particularly effective in persuading viewers, as they engage multiple cognitive processes simultaneously.

The cognitive processing of multimodal arguments involves integrating information from different modalities, evaluating the strength and relevance of this information, and drawing conclusions based on this integrated assessment. This process can lead to more robust and memorable persuasive effects compared to single-modality arguments.

The role of cognitive fluency in multimodal advertising texts is another area that merits attention. Cognitive fluency refers to the ease with which information is processed [Alter A. L. & Oppenheimer D. M., 2009: 219]. In multimodal advertising, the integration of verbal and nonverbal elements can significantly impact cognitive fluency.

For example, well-integrated verbal and visual elements can enhance cognitive fluency by providing multiple, complementary routes to understanding the advertising message. Conversely, poorly integrated or conflicting elements can reduce cognitive fluency, potentially leading to confusion or disengagement.

The cognitive processing of fluent multimodal advertisements involves smoother integration of information from different modalities, potentially leading to more positive evaluations and increased persuasiveness. However, it's worth noting that in some cases, introducing a degree of disfluency can be strategically beneficial, as it can prompt deeper processing and engagement with the advertisement [Alter & Oppenheimer, 2009: 225].

Another important cognitive feature in multimodal advertising texts is the concept of conceptual blending. Developed by Fauconnier and Turner, conceptual blending theory describes how elements from different mental spaces are combined to create new, emergent meanings [Fauconnier G. & Turner M., 2002: 40]. In multimodal advertising, verbal and nonverbal elements can serve as inputs for conceptual blends, creating rich and complex meanings that go beyond the sum of their parts.

For example, an advertisement might combine verbal metaphors with visual imagery to create a conceptual blend that conveys abstract product qualities in a concrete, memorable way. The cognitive processing of such blends involves mapping elements from different input spaces, identifying relevant connections, and constructing a blended space that generates new inferences and meanings [Forceville, 2020: 87].

The use of multimodal rhetoric in advertising texts is another area that engages multiple cognitive processes. While traditional rhetoric has focused on verbal persuasion, multimodal rhetoric extends these principles to include visual and other nonverbal elements [Scott, 1994: 255]. In multimodal advertising, rhetorical figures such as metaphor, metonymy, and irony can be realized through the interaction of verbal and nonverbal components.

For instance, a visual metaphor might be reinforced or elaborated by accompanying text, creating a multimodal rhetorical figure that engages both visual and verbal processing. The cognitive processing of such multimodal rhetoric involves recognizing the figurative nature of the message, mapping relevant conceptual domains, and deriving implicatures based on this mapping [McQuarrie & Phillips, 2005: 10].

The role of multimodal coherence in advertising texts is another important cognitive consideration. Coherence refers to the logical and semantic connections that make a text meaningful and unified [van Leeuwen, 2005: 179]. In multimodal advertising, coherence must be achieved not only within each modality but also across modalities.

For example, the visual layout of an advertisement should support and enhance the logical flow of the verbal text, while the verbal elements should provide context and elaboration for the visual components. The cognitive processing of multimodal coherence involves integrating information. Certainly, I'll continue the discussion on cognitive features of advertising texts with verbal and nonverbal components.

The cognitive processing of multimodal coherence involves integrating information from different modalities and constructing a unified mental representation of the advertisement's message. This process draws on both bottom-up perceptual processing and top-down conceptual knowledge, as viewers interpret the relationships between verbal and nonverbal elements [Schnitz & Bannert, 2003: 145].

Another significant cognitive feature in multimodal advertising is the concept of perceptual fluency. This refers to the ease with which individuals can process and interpret sensory information [Reber et al., 2004: 365]. In multimodal advertisements, perceptual fluency can be enhanced by ensuring that verbal and nonverbal elements are easily recognizable, well-integrated, and consistent with viewers' expectations and prior knowledge.

High perceptual fluency in multimodal advertisements can lead to positive affective responses and more favorable evaluations of the advertised product or brand. This is because fluent processing is often associated with familiarity and truthfulness, which can unconsciously influence consumer judgments [Alter & Oppenheimer, 2009: 227].

The role of working memory in processing multimodal advertising texts is also crucial. Working memory has limited capacity, and multimodal advertisements must be designed to optimize the use of this cognitive resource [Baddeley, 2003: 830]. By distributing information across verbal and nonverbal channels, advertisers can potentially reduce cognitive load and enhance information processing and retention.

However, it's important to note that poorly designed multimodal advertisements can overload working memory, leading to reduced comprehension and effectiveness. This underscores the importance of careful integration of verbal and nonverbal elements to support, rather than hinder, cognitive processing [Sweller et al., 2011: 57].

The concept of multimodal literacy is becoming increasingly relevant in the context of advertising texts. As consumers are exposed to more sophisticated multimodal messages, they develop skills in interpreting and critically analyzing these complex texts [Kress & van Leeuwen, 2006: 17]. This evolving literacy influences how consumers cognitively process and respond to multimodal advertisements.

For instance, consumers with high multimodal literacy may be more adept at recognizing subtle visual-verbal interactions, understanding complex metaphorical blends, and critically evaluating the persuasive strategies employed in multimodal advertisements. This highlights the dynamic nature of cognitive processing in advertising, as it evolves in response to changing media landscapes and consumer competencies [Forceville, 2020: 112].

The role of cultural cognition in processing multimodal advertising texts is another important consideration. Cultural background significantly influences how individuals interpret and respond to visual and verbal cues [Nisbett & Masuda, 2003: 11163]. In multimodal advertising, this means that the same combination of verbal and nonverbal elements may be processed and understood differently across cultural contexts.

For example, color symbolism, visual metaphors, and even the balance between text and image can have different cognitive and emotional impacts depending on the cultural background of the viewer. Effective multimodal advertising must therefore consider these cultural cognitive factors to ensure that the intended message is accurately conveyed and processed across diverse audiences [Zaltman, 2003: 78].

The concept of multimodal framing is also relevant to the cognitive processing of advertising texts. Framing refers to the selection and emphasis of certain aspects of a perceived reality, promoting a particular interpretation or evaluation [Entman, 1993: 52]. In multimodal

advertising, framing can occur through the interplay of verbal and nonverbal elements, guiding viewers' attention and shaping their interpretation of the message.

For instance, the visual composition of an advertisement can direct attention to specific elements, while the verbal text provides a framework for interpreting these visual cues. The cognitive processing of such multimodal frames involves selectively attending to certain aspects of the advertisement, activating relevant schemas, and constructing meaning based on the interaction between verbal and nonverbal cues [Kress & van Leeuwen, 2006: 177].

The role of embodied cognition in processing multimodal advertising texts is an emerging area of research. Embodied cognition posits that cognitive processes are deeply rooted in the body's interactions with the world [Barsalou, 2008: 618]. In the context of multimodal advertising, this suggests that the physical and sensory experiences evoked by verbal and nonverbal components. This article explores the cognitive features of advertising texts that incorporate both verbal and nonverbal components. It emphasizes the complex interplay between linguistic and visual elements in creating meaning, influencing consumer perception, and enhancing message retention. Key cognitive features discussed include:

1. Multimodal coherence and integration
2. Perceptual fluency and its impact on consumer judgments
3. Working memory and cognitive load in processing multimodal advertisements
4. Multimodal literacy and its evolution in response to sophisticated advertising
5. Cultural cognition and its influence on interpreting multimodal cues
6. Multimodal framing and its role in guiding attention and interpretation
7. Embodied cognition in processing sensory-rich advertising experiences

The article highlights the importance of carefully balancing verbal and nonverbal elements to optimize cognitive processing, enhance comprehension, and increase advertising effectiveness. It also emphasizes the need for advertisers to consider cultural factors, evolving consumer competencies, and the principles of cognitive psychology when designing multimodal advertisements.

Conclusion

The exploration of cognitive features in advertising texts with verbal and nonverbal components reveals the intricate and multifaceted nature of consumer information processing. This research underscores the critical importance of understanding how the human mind interacts with multimodal advertising content, offering valuable insights for both academic study and practical application in the field of marketing communication.

Several key conclusions can be drawn:

1. Integration is crucial: The effective integration of verbal and nonverbal elements is paramount in creating cohesive and impactful advertisements. This integration must be carefully designed to enhance rather than hinder message comprehension and retention.
2. Cognitive load management: Advertisers must be mindful of the cognitive load imposed by their multimodal messages. Balancing information richness with processing ease is essential for optimal communication effectiveness.
3. Cultural cognition matters: The interpretation of multimodal advertising cues is significantly influenced by cultural factors. This highlights the need for culturally sensitive approaches in global advertising campaigns.
4. Evolving consumer competencies: As consumers become increasingly sophisticated in their multimodal literacy, advertisers must continually adapt their strategies to engage and persuade their target audiences effectively.
5. Embodied experiences: The role of embodied cognition in processing sensory-rich advertising experiences offers new avenues for creating memorable and impactful campaigns.
6. Interdisciplinary approach: The study of cognitive features in multimodal advertising texts benefits greatly from an interdisciplinary approach, drawing insights from linguistics, psychology, neuroscience, and cultural studies.

As the advertising landscape continues to evolve with new technologies and media platforms, understanding these cognitive features becomes increasingly important. Future research should focus on how emerging technologies, such as augmented and virtual reality, impact the cognitive processing of multimodal advertising content. Additionally, longitudinal studies examining the long-term effects of exposure to sophisticated multimodal advertisements on consumer cognition and behavior would provide valuable insights for both theorists and practitioners in the field.

In conclusion, the study of cognitive features in advertising texts with verbal and nonverbal components offers a rich and dynamic field of inquiry. It not only enhances our understanding of human information processing but also provides practical guidelines for creating more effective and engaging advertisements in an increasingly complex media environment.

References:

1. Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13(3), 219-235.
2. Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829-839.
3. Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.
4. Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58.
5. Forceville, C. (2020). *Visual and Multimodal Communication: Applying the Relevance Principle*. Oxford University Press.
6. Kress, G., & van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design* (2nd ed.). Routledge.
7. Nisbett, R. E., & Masuda, T. (2003). Culture and point of view. *Proceedings of the National Academy of Sciences*, 100(19), 11163-11170.
8. Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4), 364-382.
9. Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141-156.
10. Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer.
11. Zaltman, G. (2003). *How Customers Think: Essential Insights into the Mind of the Market*. Harvard Business School Press.
12. Additional References:
13. Forceville, C. (2020). *Visual and Multimodal Communication: Applying the Relevance Principle*. Oxford University Press.
14. Kress, G., & van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design* (2nd ed.). Routledge.
15. McQuarrie, E. F., & Phillips, B. J. (2005). Indirect Persuasion in Advertising: How Consumers Process Metaphors Presented in Pictures and Words. *Journal of Advertising*, 34(2), 7-20.
16. Paivio, A. (2014). *Mind and Its Evolution: A Dual Coding Theoretical Approach*. Psychology Press.
17. Schnotz, W., & Bannert, M. (2003). Construction and Interference in Learning from Multiple Representation. *Learning and Instruction*, 13(2), 141-156.
18. Scott, L. M. (1994). Images in Advertising: The Need for a Theory of Visual Rhetoric. *Journal of Consumer Research*, 21(2), 252-273.
19. Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition* (2nd ed.). Blackwell Publishers.

20. van Leeuwen, T. (2005). Introducing Social Semiotics. Routledge.
21. Wedel, M., & Pieters, R. (2008). Eye Tracking for Visual Marketing. Now Publishers Inc.
22. Zaltman, G. (2003). How Customers Think: Essential Insights into the Mind of the Market. Harvard Business School Press.

SEMANTIK VALENTLIK TUSHUNCHASINING YUZAGA KELISH USULLARI

Babajanova Lobar

Phd student, Urgench State University

Annotation: this article examines the concept of semantic valency, its nature, and its differences from syntactic valency. The study is based on the theories of Yu. D. Apresyan, Ch. Fillmore, and R. Rasulov, analyzing the semantic valency of verbs and nouns and their manifestation in English, Russian, and Uzbek. The article explores thematic roles such as agent, patient, and instrument, highlighting their function in word combinations and discussing the significance of semantic valency in contemporary linguistic research. Additionally, the paper emphasizes the role of semantic valency theory in artificial intelligence, natural language processing (NLP), machine translation, and the development of semantic databases such as FrameNet. The study highlights the insufficient research on semantic valency in Uzbek linguistics, particularly regarding nouns and other word classes. Furthermore, the research discusses the importance of semantic valency in lexicography, syntactic analysis, and automated text processing, focusing on how verbs and their valency potential shape sentence structure. The findings contribute to linguistic modeling, machine translation, NLP advancements, and improving human-computer interaction efficiency.

Keywords: semantic valency, syntactic valency, artificial intelligence, natural language processing (NLP), FrameNet, thematic roles, machine translation, semantik-sintaktik guruh.

Аннотация: в данной статье рассматривается понятие семантической валентности, ее природа и отличия от синтаксической валентности. Исследование основано на теориях Ю. Д. Апресяна, Ч. Филлмора и Р. Расулова, анализируется семантическая валентность глаголов и существительных и их проявление в английском, русском и узбекском языках. В статье рассматриваются тематические роли, такие как агент, пациенс и инструмент, подчеркивается их функция в словосочетаниях и значимость семантической валентности в современных лингвистических исследованиях. Кроме того, подчеркивается роль теории семантической валентности в области искусственного интеллекта, обработки естественного языка (NLP), машинного перевода и разработки семантических баз данных, таких как FrameNet. В исследовании акцентируется недостаточная изученность семантической валентности в узбекском языкознании, особенно в отношении существительных и других частей речи. Также в статье обсуждается важность семантической валентности в лексикографии, синтаксическом анализе и автоматизированной обработке текста, с акцентом на то, как глаголы и их валентные возможности формируют структуру предложения. Полученные результаты способствуют развитию лингвистического моделирования, машинного перевода, NLP и повышению эффективности взаимодействия человека и компьютера.

Ключевые слова: семантическая валентность, синтаксическая валентность, искусственный интеллект, обработка естественного языка (NLP), FrameNet, тематические роли, машинный перевод, семантико-синтаксическая группа.

Semantika haqida gap ketganda, Moskva semantik maktabining eng yorqin vakillaridan biri bo'lmish Yu. D. Apresyanni eslamalik mumkin emas. Yu. D. Apresyan nazariy semantika va leksikografiya nuqtai nazaridan semantik valentlik masalasiga jiddiy e'tibor qaratdi. Olim o'zining "Leksik semantika" asarida "semantik valentliklar so'zning leksik ma'nosidan bevosita kelib chiqishini, uni boshqalardan farq qiluvchi o'ziga xos leksik birlik" ekanligini izohlaydi. U semantik valentlik (aktantlar) bilan bog'liqlikning boshqa turlarini bir-biridan farqlash kerakligini uqtiradi, ba'zi so'zlarning valentlik imkoniyatlari kam bo'lishini va ularning morfologik ifodasi ko'p hollarda, idiomatik bo'lishini hamda bu valentlik qaysi so'zga tegishli ekanligiga ham bog'liq

ekanligini ta'kidlaydi. Yu.D.Apresyan semantik valentliklarni boshqa bog'liqlik turlari bilan o'zaro taqqoslab, boshqa bog'liqliklarning morfologik tuzilishi ko'proq yoki kamroq ekanligi, ko'pincha ma'no bilan emas, balki tegishli tobe bog'lanish mazmuni bilan belgilanishini aytadi. Lingvist lug'atlar tuzish va lug'atlarni o'rganish jarayonida valentlik nazariyasining asosiy tushuncha va qoidalarini amaliyotga tadbiiq qilgan holda, Yu.D. Apresyan bundan keying tuziladigan lug'atlar va ma'lumotnomalarda so'zning valentligi to'g'risidagi ma'lumotlarni ham kiritish zarurligini ta'kidlab, buni so'zning "sintaktik zonasi" deb izohlaydi [Apresyan, 1995; 121].

Semantik valentlik fe'lining o'ziga xos xususiyatlarini ko'rsatadi va u grammatikadagi sintaktik valentlik tushunchasidan tubdan farqlanadi, boisi shundaki, sintaktik valentlik ko'proq grammatik tizimga asoslangan bo'lsa, semantik valentlik esa so'zning ma'nosiga va kontekstga bog'liq bo'ladi. Misol uchun, ba'zi bir fe'llar nisbatan ko'proq argumentlarni talab qilsa ya'ni bir emas bir nechta so'zlarga bog'lanib, o'zaro valentlik munosabatlariga kirishsa, boshqalari esa nisbatan kamroq, faqat bitta argument talab qilishi yokida umuman talab qilmasligi ham kuzatiladi. Ushbu tushuncha, asosan, fe'llarning gapdagi boshqa elementlar bilan qanday munosabatlarga kirishishi va so'zlar bog'langanda qanday semantik talablarni o'z ichiga olishini yanada chuqurroq anglashga yordam beradi.

Yana bir faktni ham ta'kidlab o'tish kerakki, semantik valentlik hodisasi o'zbek tilshunosligida ham jahon tilshunosligida ham asosan fe'l so'z turkumiga mansub bo'lgan so'zlar misolida o'rganilgan. Xususan, ismlar, shaxs nomlari, atoqli otlarning semantik valentlik munosabatlari deyarli o'rganilmagan desak ham bo'ladi. Xattoki, ushbu yo'nalishda 20000 dan ortiq tadqiqotlar olib borgan olim H.Shumaxer ishlarida ham semantik valentlikka aloqador ma'lumotlar juda kam keltirilgan [Sommerfeld, Herbert, 1980; 432]. O'zbek tilshunosligiga kelsak, ushbu masala, asosan, ot so'z turkumi aspektida qisman o'rganilgan va yoritib berilgan [Abdullayev, 1971; 36-41]

Avvalo, semantik valentlik tushunchasini anglashdan oldin **sintaktik valentlik** tushunchasiga nazar tashlamoq zarur bo'ladi. Chunki, bu tushunchalar bir- biri bilan uzviy bog'liq atamalar hisoblanadi. Xullas, fe'lining ot so'z turkumidagi so'z shakllari bilan bog'lanish munosabatlariga sintaktik valentlik shakllari deyiladi. Bu yerda ravish+fe'l (yaxshi o'qimoq, chiroyli gapirmoq) yoki ot+ko'makchi/bog'lovchi/yuklama+fe'l (kompyuter orqali tarjima qilmoq, do'sti bilan suhbatlashmoq) kabi konstruksiyalar nazarda tutilmaydi. O'zbek tilida 6 ta kelishiklar mavjud bo'lib, bularga bosh, qaratqich, jo'nalish, tushum, o'rin-payt va chiqish kelishiklari kiradi. Ot+ fe'l konstruksiyasi yuqoridagi kelishiklardan faqat 4 tasi bilan munosabatga kirisha oladi: masalan, sotib olmoq fe'li (to buy) jo'nalish kelishigidagi ot turkumidagi so'z bilan bog'lansa- qizimga (kiyimni) sotib olmoq, tushum kelishigidagi so'z bilan-daftarni sotib olmoq, o'rin -payt- shaharda sotib olmoq, chiqish kelishigi bilan esa- do'kondan sotib olmoq kabi so'z birikmalari sintaktik bog'lanishli shakllarga misol bo'ladi. [Abdullayev, Seyidov, Hasanov, 1972; 475]

Fe'l turkumidagi so'zlarni faqat sintaktik munosabatlariga qarab ajratish har doim ham samarali usul hisoblanmaydi va grammatik qoidalar asosida ish yuritish ham doimo yetarlicha natija bermasligi mumkin. Biz bunday holatlarda so'zlarning semantik tomoniga alohida ahamiyat qaratishimiz zarur bo'ladi, ya'ni sintaktik valentlikning tegishli komponenti orqali bo'g'langan fe'l so'z turkumiga kiruvchi elementlar semantik valentlik shakllariga misol bo'la oladi. Keling, ushbu misolni birgalikda tahlil qilaylik:

Bu kitobda shaxmat o'ynash qoidalari tasvirlangan.

(Rules of playing chess have been explained in this book.)

Kitobda so'zi (in the book) o'rin-payt kelishigi qo'shimchasini olgan (kitob+da), va o'ynash (to play) ot kesimi sintaktik munosabatlar asosida shaxmat so'zi bilan bog'langan va 0110 valentlikka ega. Fe'lining valentligini ifodalashdagi ushbu 4 ta raqamning

- 1- si jo'nalish kelishigini
- 2- si tushum kelishigini

3- si o‘rin-payt kelishigini

4- si chiqish kelishigidagi so‘zlar bilan bog‘lanish munosabatlarini ifodalashini hisobga oladigan bo‘lsak, **o‘ynash (to play) fe’li** jo‘nalish kelishigida – 0, tushum kelishigida- 1 (chunki o‘yinchoqni o‘ynamoq), o‘rin-payt kelishigida- 1 (chunki maydonda o‘ynamoq) chiqish kelishigida – 0 valentlik ko‘rsatkichlariga ega bo‘ladi. Shu asosda shaxmat o‘ynamoq (to play a chess) so‘z birikmasini kitobda so‘zi bilan semantik valentlik munosabatlarini tahlil qiladigan bo‘lsak, ular o‘zaro munosabatlarga kirishmaydi. Aksincha, kitobda (in the book) so‘zi tasvirlangan (have been explained) fe’li bilan valentlik munosabatlari hosil qiladi, kitobda tasvirlangan kabi (o‘rin -payt kelishigidagi qo‘shimchani olgan va 0010 valentlik ko‘rsatkichiga ega) [(Fatullayev, 2007)]

Ba’zi bir manbalarda fe’l bog‘lana olishi mumkin bo‘lgan so‘zlar qatori **semantik valentlik** ekanligi ta’kidlanadi. Ya’ni predikatlar(kesim) orqali ifodalangan munosabatlar ikki yoki undan ortiq argumentlarni o‘z ichiga olishi mumkin. Argumentlar, fe’l bog‘lana oladigan elementlar ya’ni so‘zlar, ularning soni qanchalik ko‘p yoki kam bo‘lishi fe’lning valentlik munosabatlarini belgilaydi. Masalan, ingliz tilidagi *to laugh (kulmoq)* fe’lini olsak, faqat bitta element bilan munosabatga kirisha oladi:

Laugh: {Malika, Mubina, Yasmina}

1. Mubina laughed. (Mubina kuldi) gapida fe’l faqat bitta argument oldi.

Keyingi misol *invite (taklif qilmoq)* so‘zini oladigan bo‘lsak, bu yerda ikkita guruh o‘rtasida munosabatlar o‘rnatiladi, ya’ni taklif qiluvchilar guruhi- mezbolar va taklif qilinuvchilar guruhi- mehmonlar. Bu yerda, Madinani mezbolar va Yasminani mehmon sifatida olsak, ‘Madina invited Yasmina’ (Madina Yasminani taklif qildi) sodda gapi yuzaga keladi, ushbu gapdagi fe’l bo‘lsa ikkita argument bilan valentlik munosabatlariga kirishadi. Agar ushbu gapga yana bitta argument qo‘shadigan bo‘lsak, ‘Madina invited Yasmina to her birthday party’ (Madina Yasminani tug‘ilgan kuniga taklif qildi) bo‘lsa, *invite* fe’li birdaniga 3 ta argument olgan bo‘ladi. Bu yerda fe’l ham vositali ham vositasiz to‘ldiruvchi bilan bog‘langan⁹².

Charls Filmor ham semantik valentlik tushunchasini tadqiq qilgan taniqli olimlardan biri bo‘lib, asosan, uning izlanishlari *Case Grammar* va *Frame semantics* nazariyalariga asoslangan. Olim o‘zining ko‘p yillik izlanishlari natijasida 1968-yilda o‘zining *Case Grammar* (Kelishiklar grammatikasi) nazariyasiga asos soladi va fe’llarning valentligi sintaktik emas, balki semantik asosga ega ekanligini isbotlaydi. Uning nazariyasiga ko‘ra, har bir fe’l o‘zining semantik argumentlari yoki aktantlari bilan uzviy bog‘langan bo‘lib, bu argumentlar turli xil vositalar orqali tasniflanadi. Olim semantik valentlikni yanada yaxshiroq anglatish uchun gapdagi so‘zlarning tematik rollari ya’ni vazifalarini guruhlariga ajratadi:

Agens- bu ish-harakatni bajaruvchi shaxs ya’ni harakatning egasi (masalan Mubina)

Patsiyent (Patient) – ish- harakat ta’sir qiluvchi narsa (masalan, daftar bo‘lishi mumkin)

Vosita (Instrument) – ish- harakat amalga oshiriladigan vosita (qalam bilan chizmoq singari)

Maqsad (Goal) – ish- harakatning yo‘nalgan obyekti ya’ni boshlanadigan va nihoyasiga yetadigan joyi (uyga bormoq kabi)

Manba (Source) – ish- harakat boshlanadigan nuqta (do‘kondan olmoq)

Tajriba qiluvchi (Experiencer) – ma’lum bir ruhiy holatni boshdan kechiruvchi shaxs yoki narsa (U xursand bo‘ldi.)

Yuqorida keltirilgan nazariyaga asoslangan holda, bitta gapni tahlil qilib ko‘raylik:

Mubina qalam bilan daftarga chizdi.

Ushbu berilgan gapda:

Agent- Mubina

Patsiyent- daftar

Vosita- qalam

bo‘ladi, *chizmoq* fe’lining semantik valentligi 3 ta argument ya’ni agent, patsiyent va vosita talab qiladi.

⁹² Some basic linguistic relations. E-Book. Chapter 3. Semantic valency.

Olim keyinchalik ilgari surgan Frame Semantics (Semantik ma'nolar tizimi) bo'lsa ushbu nazariyani yanada boyitdi. Unga ko'ra, har bir fe'l bir freym ya'ni ma'lum bir ma'no doirasida ishlatiladi va shuningdek, fe'lning valentligi faqatgina sintaktik valentlik bilangina cheklanib qolmay, semantikaga ham uzviy aloqador bo'ladi. Masalan, *Yasmina dorixonadan dori sotib oldi* gapini Case grammar nuqtayi nazaridan tahlil qilsak, quyidagi tematik rollar ajratiladi ya'ni fe'l o'ziga quyidagi argumentlarni talab qilishi mumkin: Xaridor (Buyer), bu yerda sotuvchi (Seller) ham bo'ladi, albatta, mahsulot (Goods), Narx (Price) va hokazo.

Gapni Frame Semantics jihatidan ko'rib chiqadigan bo'lsak, *sotib olmoq* fe'li *savdo freymiga* aloqador so'z hisoblanadi. Bu yerda:

Yasmina- xaridor ya'ni ish-harakatni bajaruvchi shaxs;

Dorixona- sotuvchi yoki manba ya'ni ish-harakat boshlanadigan nuqta;

Dori- mahsulot.

Filmorning semantik valentlik nazariyasi **Framenet** loyihasiga asos bo'lib, hozirda su'niy intellektning bir qator sohalarida, mashina tarjimai va **NLP** (Natural Language Processing) sohalarida, shuningdek zamonaviy kompyuter lingvistikasida keng qo'llanilmoqda.

Ma'noviy ya'ni semantik va sintaktik, lisoniy va nutqiy valentlik tushunchalarini olimlar uzoq yillar davomida o'rganib kelmoqdalar. Mazkur masala yechimi yuzasidan o'zbek lingvist olimi R.Rasulov va uning shogirdlari jiddiy ish olib borishgan. Olim ushbu masalani imkoniyat-voqelik, lison-nutq, sabab-oqibat dialektikasi aspektlari yuzasidan o'rganib, lisoniy ya'ni semantik valentlik hodisasini imkoniyat, sintaktik yoki nutqiy valentlikni bo'lsa voqelik sifatida izohlaydi. Shuningdek, olim o'z izlanishlarida valentlik tushunchasining ham sintaksis, ham semantikaga aloqador hodisa ekanligini dialektik nuqtai nazardan ochib beradi. [Rasulov, 1989; 142]

Agarda valentlik hodisasini lisoniy birliklar va til qurilishi tavsifiga substansial yondashish nuqtai nazaridan ko'rib chiqadigan bo'lsak, muayyan bir so'zning ma'lum bir ma'noviy guruh so'zlari bilan meyoriy tarzda birikish qobiliyati so'z valentligi hisoblanadi. Shundan kelib chiqadigan bo'lsak, leksik-semantik valentlik qobiliyat yoki imkoniyatdir. Shuningdek, semantik valentlik hodisasi ma'noviy, lisoniy yoki leksik-semantik valentlik nomlari bilan ham yuritiladi. Boshqacha qilib tushuntiradigan bo'lsak, so'zlarning nutqda o'zaro birikish va munosabatga kirishish imkoniyatlari semantik yoki semantik-sintaktik valentlik deb ataladi. Lisoniy ya'ni semantik valentlik so'zning semantik strukturasida jamlangan yashirin imkoniyat, mohiyat, zaruriyat, umumiylik bo'lsa, nutqiy valentlik aynan lisoniy valentlikning ro'yobga chiqishidir. Misol uchun, *bormoq* leksemasi o'z ma'noviy xususiyatlariga ko'ra bir qancha so'zlar bilan ifodalanishni talab qiladi – borish ish-harakati kimdir tomonidan, ma'lum bir maqsadda, ma'lum bir makon, zamon, tarz, sharoitda, ma'lum bir vositalar orqali bajariladi. *Bormoq* leksemasini ma'noviy xususiyatlarini tahlil qiladigan bo'lsak, bir qancha ma'nolarni ifodalaydigan va bir qator so'roqlarga javob bo'la oladigan so'zlar bilan birika olish qobiliyatiga ega bo'ladi (kim bilan bormoq, nima uchun bormoq, qayerga bormoq singari). Bu so'zning lisoniy ya'ni ma'noviy qobiliyati hisoblanadi. Nutqda, qo'llanilishda mana shu imkoniyat muayyan moddiy shaklda voqelanadi.

Shuningdek, semantik valentlik munosabatlariga qarab so'zlarni o'ziga xos guruhlariga ajratish metodi tilshunoslikda juda ham qadimiy tamoyillardan biri hisoblanib, tarixiy ildizlari uzoq yunon hamda hind tilshunoslik an'analarga borib taqaladi. Biroq bunday bo'linish ya'ni guruhlash, asosan, fe'l so'z turkumiga mansub leksik birliklarni o'timli va o'timsiz kabi ikkita katta turlarga ajratishdan nariga bormagan. Undan keyingi davrlar tilshunosligida bo'lsa leksikani tizim sifatida talqin etish zarurati asosida so'zlarni ma'noviy umumiyliklar asosida har xil uyalarga ya'ni leksik-semantik guruhlar, (lug'aviy-ma'noviy guruhlar) to'dalar va maydonlarga birlashtirgan tarzda o'rganish avj olgan. Shu jarayonda muhim bo'lgan xulosalardan biri shuki, valentlik har xil qobiliyatga ega hodisa, hatto bir vaqtning o'zida bir-biriga o'zaro zid bo'lgan so'zlar ham bir lug'aviy-ma'noviy guruh, to'da, yoki maydon kabi to'plamlarga mansub bo'lishi mumkin ekan. Keyinchalik, ushbu tadqiqotlar **semantik-sintaktik guruh** tushunchasini chuqurroq o'rganishga va tahlil qilishga sabab bo'ldi.

Semantik-sintaktik guruh (qisqacha **SSG**) tushunchasi tilshunoslikda nisbatan yangi tushunchalardan biri hisoblanib, valentlik qobiliyatlari aynan bir xil va ma'lum bir o'ta umumiy ma'no orqali birlashish xususiyatiga ega bo'lgan leksik birliklar guruhi tushuniladi. O'zbek tilshunosligiga ushbu terminni A.A.Salkalamanidze olib kirganligi ta'kidlanadi. Olim fe'llarning o'timli va o'timsizga ajratilish an'anasini davom ettirgan holatda, o'zbek tilidagi o'timsiz fe'llarni medial va nomedial, o'timlilarni bo'lsa refleksiv va norefleksiv, natijali va natijasiz faoliyat fe'llari, ularning o'zini yana bir qancha jami 9ta kichik ichki guruhchalarga ajratadi va tilshunoslikda ajratiladigan fe'llarning 56 turdagi barcha ma'noviy guruhlarini ushbu SSGlarga birlashtiradi. Shuningdek, ajratilgan SSGlarning har biri uchun xos bo'lgan minimal modellarning o'zaro munosabatini, fe'llarning zaruriy valentligini miqdoriga tayanib bir valentli, ikki valentli, to'rt valentli kabi turlarga ajratadi [Salkalamanidze, 1987; 129].

SSG tushunchasi tilshunoslikda, asosan, fe'l so'z turkumiga doir leksik birliklar misolida o'rganilgan bo'lib, qator o'zbek olimlari A.Salkalamanidze ajratgan SSGlarga tayangan holda, fe'l nisbatlari ma'no va vazifalarini, ularning so'z o'zgartirish va so'z yasashga oid munosabatlarini izohlaganlar. [Muhammedova, 2007; 402], [Rasulov, 1991;177], [Sultonsaidova, 1994; 57]

Shuni ham alohida ta'kidlash zarurki, qadim yunon tilshunoslik maktablaridan to hozirgi kunga qadar semantik va sintaktik xususiyatlariga qarab ajratilib kelinayotgan barcha so'z turkumlari, aslida o'ziga xos SSG lar hisoblanadi. Har bir so'z turkumining o'ziga xos umumiy birikuvchanlik xususiyati ham mavjud bo'lib, ushbu jihat turkum ma'nosi bilan chamcharbas bog'langan ya'ni so'zning muayyan ma'nosi bilan aloqador bo'lmaydi. Masalan, ot so'z turkumiga mansub istalgan bir leksik birlik belgi, miqdor, xususiyat, egalik qo'shimchalari bilan aloqaga kirishgan taqdirda, qaratqichli-qaratuvchi neytral valentlikka ega bo'ladi. Asosan, aniqlovchi vazifasida keluvchi ushbu elementlar so'zning fakultativ va zaruriy valentliklari, so'zning semantik jihatlari bilan bevosita aloqador bo'lmaydi. [Hojiyev, 2010; 88-98]

Shuningdek, so'z semantik valentligining muayyan ko'rinishi bu **fe'l valentligi** hisoblanadi. Chunonchi, valentlik nazariyasi dastlab fe'l so'z turkumining gap tarkibida qo'llanishining mahsuli o'laroq, fe'lning boshqaruv jihatlarini tahlil qilish mobaynida shakllantiriladi [Qo'chqortoyev, 72]. Akad. V.Vinogradov ta'kidlab o'tganidek, "...boshqa barcha so'z turkumi kategoriyalariga nisbatan qiyoslaydigan bo'lsak, eng konstruktiv turkum fe'ldir. Fe'lli birikmalar otli so'z birikmalari va gaplarga kuchli ta'sir o'tkazish kuchiga ega ekaligi bilan ham ajralib turadi" [Vinogradov, 1972; 337] . Shunday ekan, fe'lning eng muhim jihatlaridan biri, u gapning semantik strukturaviy markazida asosiy va muhim joyni band etadi [Vasilyev, 1981; 35] . Darvoqe, L.Tener tobelik grammatikasida ham gap tuzilishini tahlil qilish jarayonida gapning strukturaviy markazi sifatida fe'l so'z turkumiga asoslanadi. Demak, yuqorida keltirilgan fikrlardan kelib chiqadigan bo'lsak, gapning semantic valentligi fe'lning semantic valentligi asosida quriladi va unga tayanadi. Shuning uchun ham biz fe'lni asosiy xususiyati sifatida uni valentlik tashuvchi ekanligini ta'kidlaymiz. [Kibardina, 1982; 24]

Fe'l yoki predikatning gap konstruksiyasining tuzilishini belgilashi uning potensial semantik imkoniyati hisoblanib, uni boshqa so'z turkumlaridan ajratib turuvchi eng asosiy jihatlaridan biri sanaladi. S.Kasnelsonning fikriga ko'ra, "Mazmun planida fe'l predikat faqat muayyan leksik ma'no ifodalabgina qolmasdan, balki u bir vaqtning o'zida bo'lajak gapning maketini ham ifoda etadi. [Kanselson, 1972; 88] . Yuqorida aytilgan nazariyalardan kelib chiqib fikrimizni xulosalaydigan bo'lsak, fe'l faqat harakat yoki holatni emas, balki muayyan situatsiyani ham bir butun holda anglata olish xususiyatiga ega bo'lib, u shu harakat, holat, situatsiyaning hamma aktantlarini ya'ni qatnashuvchilarini ham birgalikda ifodalash qobiliyatiga ham ega sanaladi. Fe'l so'z turkumining aynan ana ushbu jihatiga uning semantik valentligi sifatida qaraladi (Dolinina, 68) . Semantik valentlikka ega bo'lgan fe'l nutq, harakat, holat, eshitish, ko'rish bilan bog'liq jarayonlarni ham, ushbu jarayonlarning obligator aktantlari ya'ni qatnashuvchilari o'rtasidagi o'zaro munosabatlarni ham o'zida aks ettira oladi.

Fikrimizni xulosalaydigan bo'lsak, semantik valentlik nazariyasi tilshunoslikda muhim o'rin tutishining yana bir omillaridan biri shundaki, u gapdagi so'zlararo o'zaro bog'lanishlarni

yanada chuqurroq tushunishga ko'mak beradi, hamda sintaksis va semantika o'rtasidagi aloqadorliklarni izohlashda muhim ahamiyatga ega. Ushbu nazariya tarjima nazariyasi, avtomatik tilga ishlov berish sohasi (NLP) va lingvistikaning sintaktik tahlil kabi sohalarida ham qo'llaniladi. Bugungi kunda ushbu sohada izlanishlar va tadqiqotlar hamon davom etib kelmoqda va shuningdek, sun'iy intellektga asoslangan tillarni tushunish tizimlarida ham faol ravishda qo'llanilmoqda.

Foydalanilgan adabiyotlar:

1. Apresyan, Y. 'Избранные труды, том I. Лексическая семантика. Москва: Школа «Яз. р. культуры», 1995. -B. 121
2. Sommerfeld Karl-Ernst, Schreiber Herbert. Worterbuch zur Valenz und Distribution der Substantive. WEB Bibliographisches Institut Leipzig, 1980. – P. 432
3. Abdullayev F. So'zlar o'zaro qanday bog'lanadi? - Toshkent: Fan, 1974.-75 b.;
4. Abdullaev F., Ibroximova F. O'zbek tilida boshqaruv. Toshkent: O'zbekiston, 1982. -140 b.;
5. Abdullaev F., Ibroximova F. So'z birikmasi sintaksisining rivojlanish masalalari // O'zbek tili va adabiyoti. - Toshkent, 1972. - No2. - B. 36-41.;
6. Ibroximova F. Imennoe upravlenie v sovremennom uzbekskom literaturnom yazike: Avtoref. diss. ... kand. filol. nauk. - Tashkent, 1972.- 34s.;
7. Ibroximova F. Ot boshqaruvining ayrim masalalari // O'zbek tili va adabiyoti. - Toshkent, 1971. - No2. - B. 47-51.; Ibroximova F. Sifat boshkaruvi // O'zbek tili va adabiyoti. - Toshkent, 1971. - No6. - B. 34-37
8. Abdullayev A., Seyidov Y., Hasanov A. (1972) Modern Azerbaijani language. Baku, Maarif. – B. 475
9. Fatullayev, Abulfat. The use of syntactical and semantic valences of the verb for formal delimitation of verb word phrases. 2007.
10. Rasulov R. O'zbek tilidagi holat fe'llari va ularning obligator valentliklari. -Toshkent: Fan, 1989.- B. 142
11. Salkalamanidze A.A. Семантико-синтаксические группы глаголов в узбекском языке. Мецниереба, 1987. – B. 129
12. Muhamedova S.X. O'zbek tilidagi harakat fe'llarining semantik va valentlik xususiyatlari: Filol. fan. d-ri ... diss. Toshkent, 2007. – B. 402 .;
13. Rasulov R. Лексико-семантические группы глаголов состояния и их валентность. - Tashkent: Fan, 1991.-B. 177.;
14. Sultonsaidova S. O'zbek tilidagi grammatik kategoriyalarning o'zaro munosabati: Filol. fan. d- ri ... diss. avtoref. Toshkent, 1994. –B. 57
15. Hojiev A. O'zbek tili morfologiyasi, morfemikasi va so'z yasalishining nazariy masalalari. – Toshkent: Fan, 2010. - B. 88-98
16. Ibotova Sh.O'zbek tilida otlashgan so'zlarda grammatik valentlik. Magistrlik dissertatsiyasi. Termiz. 2023.
17. Qo'chqorto'ev I.Q. So'z ma'nosi va uning valentligi. – B. 72
18. Виноградов В.В. Русский язык. М. 1972. 337-бет
19. Васильев Л. М. Семантика русского глагола. М, 1981. 35-бет
20. Кибардина С.М. Категории субъекта и объекта и теория валентности // Категории субъекта и объекта в языках различных типов. Л. 1982. 24-бет.
21. Кацнельсон С.Д. Типология языка и речевое мышление. Л. 1972. 88-бет.
22. Долинина И.Б. Маркировка субъектно-объектных отношений у валентностных категории английского глагола // Категория субъекта и объекта в языках различных типов. 68-бет.

QO'SHMA GAPLARNI PUNKTUATSION TAHLIL QILISH ALGORITMINI ISHLAB CHIQISH

Sharipov Maqsud Siddiqovich

Urganch davlat universiteti,
Kompyuter ilmlari kafedrası dotsenti

Adinayev Xushnodbek Saylboyevich

Muhammad al-Xorazmiy nomidagi
TATU Urganch filiali o'quv-uslubiy boshqarma boshlig'i

Annotatsiya: Bu ishda biz o'zbek tili matnlari qo'shma gaplarini tahlil qilishning qoidaga asoslangan algoritmini ishlab chiqishni ko'rib chiqamiz. O'zbek tili kam resursli til hisoblangaligi sababali hozirgacha bunday algoritmlar ishlab chiqilmagan. O'zbek tili matnlaridagi qo'shma gaplarda vergul yoki nuqtali vergul belgilari to'g'ri yoki noto'g'ri qo'yilganligini aniqlash masalasini yechishning qoidaga asoslangan algoritmi ishlab chiqilgan va qoidalar bazasi yaratilgan. O'zbek tilshunosligida shart munosabatlarining ifodalanish masalalari sezilarli darajada o'rganilmagan bo'lsada, qisman o'zbek tili sintaksisida tadqiq qilingan. O'zbek tilshunosi Sh. Shukurovning e'tiroficha, hozirgi o'zbek tilida shart munosabatlari bilan bog'liq quyidagi ilmiy-nazariy muammolar o'z yechimini topishi zarur ular o'z ichiga quyidagi muammolarni oladi; shart mayli formalarining zamon kategoriyasi bilan munosabati, shart munosabatini ifodalovchi analitik formalar tavsifi, shart munosabatlarining ko'rinishi va xarakteri. Ushbu maqolada o'zbek tilshunosligida qo'shma gaplarning tasnifini ko'rib chiqamiz.

Shu ma'noda aytishimiz mumkinki, tabiiy tilni qayta ishlash (NLP) bu kompyuter fanining, ayniqsa sun'iy intellektning (AI) kichik sohasi bo'lib, u kompyuterlarga inson tilini tushunish va qayta ishlashga imkon beradi. Texnik jihatdan, NLP ning asosiy vazifasi tabiiy tildagi katta hajmdagi ma'lumotlarni tahlil qilish va qayta ishlash uchun kompyuterlarni dasturlashdir.

Kalit so'zlar. Tinish belgilari, NLP, shart gap, qo'shma gap.

Kirish

Globalizatsiya va raqamli axborotlashtirish jarayonida o'zbek tilining rivojida axborot texnologiyalari o'ta muhim o'ringa ega. Kompyuter lingvistikasining tabiiy tilni qayta ishlash (Natural Language Processing – NLP) jarayoni va tabiiy tilning kompyuter tushunadigan til (formal) shakli yaratilishi natijasida tilga oid masalalar (tahrir, tahlil, tarjima, elektron matnni ovozlashtirish, og'zaki nutqni elektron matnga aylantirish, robot bilan muloqot qilish, yirik matnni ixcham matnga aylantirish kabi) bo'yicha kompyuter dasturlari yaratilmoqda.

Gaplar turli nuqtai nazardan: tuzilish, gapirish, maqsadi, barcha bulaklarning bor-yoqligiga ko'ra turlarga ajratiladi. Gaplar tuzilishi jihatdan ikki turga bo'linadi soda va qo'shma gaplar. O'z navbatida qo'shma gaplar ham alohida turlarga bo'lib o'rganiladi. Ular quyidagilardir;

- 1) bog'langan qo'shma gaplar,
- 2) ergashgan qo'shma gaplar,
- 3) murakkab qo'shma gaplar.

Ikki yo undan ortiq teng huquqli gaplardan taslikil topgan gaplar **bog'langan qo'shma gap** deyiladi. Misollar; Hamnia quvnoqlik bilan so'zlashmoqda, faqat Arslonqul negadir hayolchan o'tirardi.

Qo'shma gaplarning ergashgan turi tarkibida ikki yo undan ortiq gaplar bo'lib, ulardan biri bosh gap, qolganlari esa tobe gaplar bodadi. Bosh gapga tobe bodgan gap **ergash gap** deb yuritiladi. Ergash gaplar bosh gapdagi birorta so'zga yoki bosh gapning hammasiga tegishli bo'lib, bosh gapdan anglashilgan fikrni turli yodlar bilan to'ldirib keladi.

Misollar;

1. Xosil yigib olingach, ikki qayta urug' ekildi.

2. G'alla mo'1 bulgani uchun, omborlar todib ketdi.

3. Yomgdr tinsa, shaxar buylab sayrga chiqamiz.

Bu misollarning birinchisidagi ergash gap **payt**, ikkinchisidagi — **sabab**, uchinchisidagi — **shart** ma'nolarini ifoda etmoqda. Ergash gaplarning vazifalariga qarab, ularni **payt**, **sabab**, **shart**, **aniqlovchi**, **tuldiruvchi ergash gap** kabi turlarga ajratiladi. Bu barcha tillarga xos hususiyatdir.

Shart ma'nosini ifodalovchi gaplar fe'lning shart mayli shakli **-sa** yoki **-ganda**, **-ganda** edi vositalari orqali bogTanadi, shuningdek, ma'noni kuchaytirish uchun gapning boshida **agar**, **agarda**, **mabodo** so'zlarini qo'llash mumkin. Masalan: Agar hayot karvon bo'sa, sarbon muallim.

Shart ma'nosi sodda gaplarda **shart**, **kerak**, **zarur**, **darkor**, **lozim** so'zlari orqali ham bildiriladi. Masalan; Hamma yo'l transport qoidalariga rioya qilishi shart.

Shart ma'nosini ifodalab kelgan gaplarga misollar;

1. Mehnat qilsang, rohat ko'rasan.
2. Shamol bo'lmasa, daraxtning uchi qimirlamaydi.
3. Agar siz romanni o'qigan bo'lsangiz, mazmunini aytib bering.
4. Bahor kelsa, kunlar isib ketadi.
5. Agar astoydil ishlamoqchi bo'lsangiz, sizga men yordam beraman.
6. To'rt faslimning yozi yo'q, sen bo'lmasang yonimda.
7. Arpa ekkan arpa oladi, bug'doy ekkan bug'doy oladi.
8. Onalarda g'am qolmas, mustahkam bo'lsa tinchlik.
9. Siz menga boshini korsatib bersangiz, davomini o'zim bajaraman.

Methodologiyasi

Qo'shma gaplarda punktuatsiya (tinish belgilarining) to'g'ri qo'llanilishi juda muhim, chunki u gapdagi mazmuniy bog'lanishni aniq ifodalaydi va o'qishni osonlashtiradi.

Quyida qo'shma gaplarda tinish belgilarining qanday qo'llanishi haqida batafsil ma'lumot va misollar keltirilgan:

1. Bog'lovchisiz qo'shma gaplarda

Bog'lovchi bo'lmagani uchun **vergul (,)** yoki **nuqtali vergul (;)** ishlatiladi:

Agar gaplar ohangda keskin ajralmasa → vergul ishlatiladi:

1. Deraza ochiq, xona salqin.
2. Quyosh botdi, qushlar uyasiga qaytdi.

Agar mazmuniy jihatdan kuchliroq bo'linish bo'lsa → nuqtali vergul ishlatiladi:

1. O'quvchilar sinfga kirdilar; ustoz darsni boshladi.

2. Bog'lovchili qo'shma gaplarda

Bu yerda bog'lovchilar ishtirok etadi, shuning uchun **tinishi** ularning o'rniga qarab o'zgaradi:

Bog'lovchi orasida kelgan bo'lsa:

Vergul qo'yilmaydi, agar gap qismlari qisqa bo'lsa:

1. Men kelaman va sen bilan boraman.

Agar gap uzun yoki ohangda tanaffus bo'lsa, vergul qo'yiladi:

1. U ko'p ishladi, lekin kutgan natijaga erisha olmadi.

Bog'lovchi boshida bo'lsa:

Ergash gapdan keyin **vergul** qo'yiladi:

1. **Agar u kelsa**, biz boshlaymiz.
2. **Chunki yomg'ir yog'moqda**, yurishga chiqa olmadik.

3. Qo'shma ergash gaplarda

Ergash gap odatda **asosiy gap bilan vergul orqali ajratiladi**, ayniqsa ergash gap **boshida** kelsa:

Ergash gap boshida:

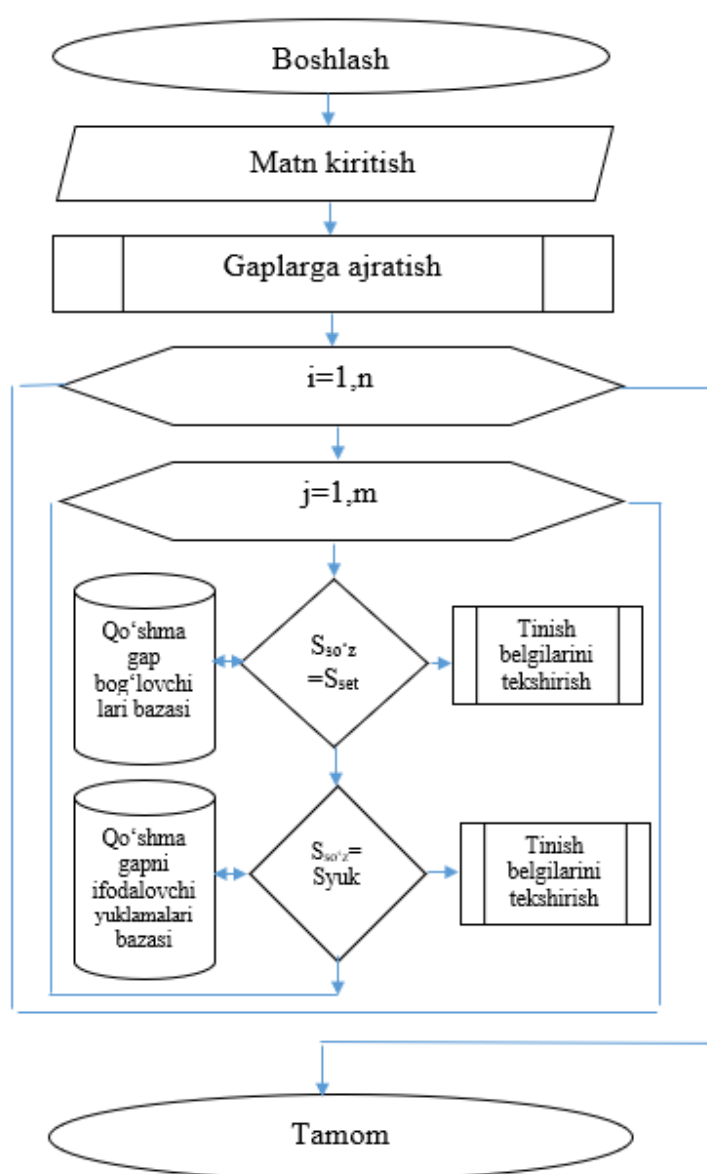
1. **Agar u ishtirok etsa**, jamoa yutadi.
2. **Garchi u yosh bo'lsa-da**, tajribasi katta.

Ergash gap keyinda kelsa, ko‘pincha vergul qo‘yilmaydi, agar ohang talab qilmasa:

1. U yutdi **chunki** juda yaxshi tayyorgarlik ko‘rgan edi.

Umumiy qoida:

Holat	Tinish belgisi
Mustaqil gaplar bog‘lovchisiz	Vergul yoki nuqtali vergul
Bog‘lovchili qo‘shma gap (uzun yoki ohangda tanaffus bo‘lsa)	Vergul
Ergash gap boshda	Verguldan keyin
Ergash gap keyinda (ko‘pincha)	Tinish belgisi qo‘yilmaydi



1-rasm

1-rasm. O‘zbek tili matnlarida qo‘shma gaplarni tinish belgilarini to‘g‘ri yoki noto‘g‘ri qo‘yilganligini tekshiruvchi algoritm. Quyida shu algoritmning python dasturlash tilidagi kodi keltirilgan.


```
import re

# Tinish belgisi noto'g'ri qo'yilishi mumkin bo'lgan holatlarni tekshiradi
def tekshir_gap(gap):
    natijalar = []

    # 1. "va", "yoki", "lekin", "ammo", "chunki" oldidan vergul kerak emas (odatda)
    noto'g'ri_vergullar = re.findall(r"\s*(va|yoki|lekin|ammo|chunki|biroq)", gap)
    if noto'g'ri_vergullar:
        natijalar.append(" Bog'lovchidan oldin vergul bo'lmasligi kerak: " + ",
        ".join(noto'g'ri_vergullar))

    # 2. Ergash gap boshida bo'lsa, vergul bo'lishi kerak
    ergash_boshida = re.match(r"^(Agar|Chunki|Garchi|Toki|Negaki|Go'yo|Go'yoki|Shuning
    uchun).*?", gap)
    if not ergash_boshida and
    re.match(r"^(Agar|Chunki|Garchi|Toki|Negaki|Go'yo|Go'yoki|Shuning uchun)", gap):
        natijalar.append(" Ergash gapdan keyin vergul qo'yilishi kerak.")

    # 3. Bog'lovchisiz qo'shma gaplarda vergul yoki nuqtali vergul bo'lishi kerak
    if 'va' not in gap and 'yoki' not in gap and 'lekin' not in gap and 'ammo' not in gap:
        qismlar = re.split(r"[.,]", gap)
        if len(kismlar := list(filter(lambda x: x.strip(), qismlar))) > 1:
            natijalar.append(" Bog'lovchisiz qo'shma gap: tinish belgisi to'g'ri qo'llanilgan.")

    if not natijalar:
        natijalar.append(" Tinish belgilarida xatolik topilmadi.")

    return natijalar

# Test uchun bir nechta gaplar
gaplar = [
    "Quyosh botdi, qushlar uyasiga qaytdi.",
    "Men keldim, va u ketdi.",
    "Agar u kelsa biz boshlaymiz.",
    "U darsga kech qoldi, chunki avtobus kelmadi.",
    "Garchi u charchagan edi lekin ishlashda davom etdi."
]

for gap in gaplar:
    print(f" Gap: {gap}")
    for x in tekshir_gap(gap):
        print(" " + x)
    print()
```

Xulosa

Hozirgi kungacha o'zbek tili matnlarida qo'shma gaplarning tinish belgilarini tahlil qilish agoritmlari deyarli ishlab chiqilmaganligi sababli ushbu ishda taklif qilingan algoritmlar dolzarb hisoblanadi. O'zbek tilida tabiiy tilni qayta ishlash masalasi bo'lgan punktuatsion tahlil (qo'shma gaplarda) qilishishni qoidaga asoslangan algoritmlari ishlab chiqildi, punktuatsion tahlil (qo'shma gaplar) uchun qoidalar bazasi yaratildi. Ishlab chiqilgan algoritmlar yordamida o'zbek

tilining 30 ta kategoriyadan iborat 1 642 860 ta soʻz va 137 848 ta gapdan iborat korpus tahlil qilinib natijalar olindi. Ushbu ishning davomi sifatida keljakda oʻzbek tili matnlarini punktuatsion (qoʻshma gaplarni) tahlil qilish masalasini sunʼiy intellekt modellaridan foydalanib yechish va yaxshiroq aniqlikdagi natijalarni olish maqsad qilingan.

Foyadalanilgan adabiyotlar roʻyxati:

1. M. A. , M. N. , A. D. Mahmudov N., *OʻZBEK TILI MEʼYORLARI (Punktuatsiya)*, vol. 1. Tashkent: Zamin nashr, 2021.
2. J. Pokrywka, "Punctuation Prediction for Polish Texts using Transformers," Feb. 2024. <https://10.48550/arXiv.2410.04621>.
3. P. Żelasko, P. Szymański, J. Mizgajski, A. Szymczak, Y. Carmiel, and N. Dehak, "Punctuation Prediction Model for Conversational Speech," Feb. 2018, pp. 2633–2637. <https://10.21437/Interspeech.2018-1096>.
4. M. S. Sharipov, H. S. Adinaev, and E. R. Kuriyozov, "Rule-Based Punctuation Algorithm for the Uzbek Language," in International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices, EDM, 2024, pp. 2410 – 2414. <https://10.1109/EDM61683.2024.10615061>.
5. M. Bayraktar, B. Say, and V. Akman, "An analysis of English punctuation: the special case of comma," International Journal of Corpus Linguistics, vol. 3, Jul. 1998, <https://10.1075/ijcl.3.1.03bay>.
6. D. Hardt, "Comma checking in Danish," 2001.
7. U. Salaev, E. Kuriyozov, and C. Gomez-Rodríguez, "SimRelUz: Similarity and Relatedness scores as a Semantic Evaluation Dataset for Uzbek Language," in 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages, SIGUL 2022 - held in conjunction with the International Conference on Language Resources and Evaluation, LREC 2022 - Proceedings, 2022, pp. 199 – 206. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85138700420&partnerID=40&md5=bf476cd74317f06577dd0548c5c60>
8. A. M. Abdurashetona and I. O. Ismailovich, "Methods of Tagging Part of Speech of Uzbek Language," in Proceedings - 6th International Conference on Computer Science and Engineering, UBMK 2021, 2021, pp. 82 – 85. doi: 10.1109/UBMK52708.2021.9558900.
9. A. M. Abdurashetona and U. Mokhiyakon, "Software Features and Linguistic Features of Uzbek Synonymizer," in Proceedings - 7th International Conference on Computer Science and Engineering, UBMK 2022, 2022, pp. 171 – 175. <https://10.1109/UBMK55850.2022.9919447>
10. B. Mengliyev, S. Shahabitdinova, S. Khamroeva, S. Gulyamova, and A. Botirova, "The morphological analysis and synthesis of word forms in the linguistic analyzer," Journal of Language and Linguistic Studies, vol. 17, no. 1, pp. 558 – 564, 2021, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103797274&partnerID=40&md5=8a4052419f721c3c734f8fe1c48984ec>
11. K. Madatov, S. Bekchanov, and J. Vivic, "Dataset of Karakalpak language stop words," Data Brief, vol. 48, 2023, doi: 10.1016/j.dib.2023.109111.
12. M. Sharipov, J. Mattiev, J. Sobirov, and R. Baltayev, 'Creating a Morphological and Syntactic Tagged Corpus for the Uzbek Language', CEUR Workshop Proceedings, vol. 3315, pp. 93–98, 2022.
13. D. Mengliyev, E. Akhmedov, V. Barakhnin, Z. Hakimov, and O. Alloyorov, "Utilizing Lexicographic Resources for Sentiment Classification in Uzbek Language," Jan. 2023, pp. 1720–1724. <https://10.1109/APEIE59731.2023.10347765>

TIL TA'LIMIDA KOMPYUTER TEXNOLOGIYALARI

Jumayeva Anorxol Chaman qizi

Mirzo Ulug'bek nomidagi

O'zbekiston Milliy universiteti 3-kurs talabasi

E-mail: anorajumayeva130304@gmail.com

Annotatsiya: Ushbu maqolada zamonaviy kompyuter va axborot-kommunikatsiya texnologiyalarining til o'rganish jarayonida tutgan o'rni va ahamiyati tahlil qilingan. Jumladan, "Kahoot", "Baraban" kabi interaktiv metod va platformalarning ta'lim samaradorligini oshirishdagi roli ochib berilgan. Shuningdek, multimedia vositalari, avtomatik tahlil va tarjima tizimlari orqali til o'rgatishda individual yondashuv va interaktivlikni ta'minlash imkoniyatlari yoritilgan. Maqola til ta'limida raqamli texnologiyalarni qo'llashning dolzarbligini asoslab, pedagogik jarayonni innovatsion yondashuv asosida tashkil etish zaruratini ko'rsatib beradi.

Kalit so'zlar: til ta'limi, kompyuter texnologiyalari, interaktiv o'rganish, onlayn platformalar, til ko'nikmalari, dasturlar, grammatika, so'z boyligi, ta'lim metodlari, individual o'rganish.

Abstract: This article analyzes the role and importance of modern computer and information and communication technologies in the language learning process. In particular, the role of interactive methods and platforms such as "Kahoot", "Baraban" in increasing the effectiveness of education is revealed. It also highlights the possibilities of ensuring an individual approach and interactivity in language teaching through multimedia tools, automatic analysis and translation systems. The article substantiates the relevance of using digital technologies in language education and shows the need to organize the pedagogical process based on an innovative approach.

Keywords: language education, computer technologies, interactive learning, online platforms, language skills, programs, grammar, vocabulary, teaching methods, individual learning.

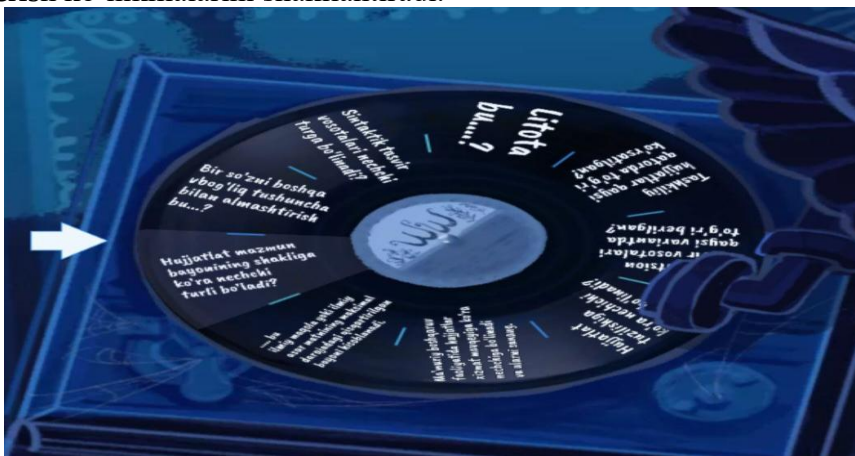
Аннотация: В статье анализируется роль и значение современных компьютерных и информационно-коммуникационных технологий в процессе изучения языка. В частности, раскрывается роль интерактивных методов и платформ, таких как «Кахут», «Барабан» в повышении эффективности образования. В нем также освещаются возможности обеспечения индивидуального подхода и интерактивности в преподавании языка с помощью мультимедийных инструментов, систем автоматического анализа и перевода. В статье обосновывается актуальность использования цифровых технологий в языковом образовании и показывается необходимость организации педагогического процесса на основе инновационного подхода.

Ключевые слова: языковое образование, компьютерные технологии, интерактивное обучение, онлайн-платформы, языковые навыки, программы, грамматика, словарный запас, методы обучения, индивидуальное обучение.

Bugungi kunda Axborot texnologiyalari ta'lim-tarbiya tizimini modernizatsiya qilishning muhim omilidir. O'zbekiston Respublikasi Prezidentining farmon va qarorlariga muvofiq, "Elektron ta'lim" milliy tarmoqlarining tashkil etilishi va ularning kontentini boyitish orqali ta'lim sifati va samaradorligini oshirishga alohida e'tibor qaratilmoqda. Xususan, O'zbek tili va adabiyoti fanini o'qitishda ilg'or pedagogik va axborot texnologiyalarini joriy etish, ilmiy-metodik tadqiqotlarni chuqurlashtirish bugungi ta'lim siyosatining dolzarb yo'nalishlaridan biridir. Bu esa zamonaviy bilimga ega, raqamli kompetensiyalarga tayangan pedagog kadrlarni tayyorlashni taqozo etadi. O'zbekiston Respublikasi Prezidenti Sh.M.Mirziyoyev 2017-yil 14- yanvar kuni Vazirlar Mahkamasining 2016- yildagi ijtimoiy-iqtisodiy rivojlantirish yakunlarini har tomonlama

tahlil qilish hamda respublika hukumatining 2017-yil uchun iqtisodiy va ijtimoiy dasturi eng muhim yo'nalishlari va ustuvor vazifalarini belgilashga bag'ishlangan kengaytirilgan majlisda qilgan ma'ruzasida: "Bolalar va yoshlarga maxsus fanlar, mamlakatimiz va jahon sivilizatsiyasi tarixini, xorijiy tillarni va zamonaviy kompyuter dasturlarini chuqur o'rgatish vazifalari hali sifatli va to'liq holda yechilgani yo'q. Yana bir muammoni hal etish o'ta muhim hisoblanadi: bu - pedagoglar va professor-o'qituvchilar tarkibining professional darajasi, ulaming maxsus bilimlaridir. Bu borada ta'lim olish, ma'naviy-ma'rifiy kamolot masalalari va haqiqiy qadriyatlarni shakllantirish jarayonlariga faol ko'mak beradigan muhitni yaratish zarur", - deb ta'kidladilar. O'zbek tili va adabiyoti ta'limida zamonaviy axborot texnologiyalarini qo'llashda talabalarga o'quv mashg'ulotlari jarayonida qo'llaniladigan pedagogik dasturiy vositalar, o'quv mashg'ulotlari uchun tayyorlanadigan resurslar va ularga qo'yiladigan talablar, bu boradagi chet el tajribasi haqida ma'lumot berish talabalarda axborot asri sharoitida pedagogik jarayonni samarali tashkil etish uchun zarur ko'nikmalarni, ijodiy va texnologik madaniyatni shakllantirishga asos bo'ladi. Ta'limda axborot texnologiyasi - bu "o'quvchi - kompyuter" o'rtasidagi muloqotdir, o'qituvchi esa uni boshqaruvchi. Zamonaviy axborotlarga boy materiallar, vositalar bilan darsni boshqarish o'qituvchi mehnatini samarali qiladi. Shuningdek, multimedia resurslaridan audio, video, animatsiya va interaktiv o'yinlarning qo'llanilishi o'quvchilarni dars jarayoniga faol jalb etadi hamda tilni amaliyotda tezroq o'zlashtirishga yordam beradi. Eshitish orqali o'rganishga asoslangan audio metodlar imkoniyati cheklangan shaxslar uchun ham samarali va qulay ta'lim muhitini yaratadi.

Til ta'limida interaktiv metodlardan foydalanish samarali natijalarni beradi. Masalan, "Baraban" metodi o'quvchilarning diqqatini oshirish va darsga jalb etish vositasi sifatida qo'llaniladi. Ushbu metoddan *wordwall* saytida foydalanish mumkin. O'qituvchi mavzuga doir savollarni platformadagi sektorlarga joylashtiradi. Baraban aylantirilganda, strelka tasodifiy savolda to'xtaydi va qatnashchilar oldindan tayyorlangan savollar yozilgan qog'ozlar (yoki kartochkalar)ni tasodifiy ravishda tanlaydi. Har bir ishtirokchi tushgan savolga og'zaki yoki yozma tarzda javob beradi. Bu metodning asosiy mohiyati — ishtirokchilar o'rtasida savol va javoblar almashinuvini tashkil etish orqali ularning mavzuga oid bilim va tushunchalarini chuqurlashtirishdan iboratdir. Bu jarayon o'rganuvchilarning o'z fikrini erkin ifoda etish, dalillash va muloqot madaniyatini rivojlantirishga yordam beradi. Metodning nomlanishi savollar "baraban ichida aralashtirilganidek" tasodifiy tanlanishi bilan bog'liq. Mazkur usul ayniqsa gumanitar fanlar, pedagogika, til o'rganish va ijtimoiy fanlar yo'nalishidagi darslarda samarali qo'llaniladi. U dars jarayonini jonlantiradi, tinglovchilarni faol ishtirok etishga undaydi va muhokama madaniyatini oshiradi. Ilmiy-tadqiqotlar shuni ko'rsatadiki, "Baraban" metodi o'quvchilarning mavzunni chuqurroq anglashiga va ijodiy fikrlash ko'nikmalarini shakllantirishga ijobiy ta'sir ko'rsatadi. Mazkur metod o'yin elementlariga asoslanib o'quvchilarda faollik, mustaqil fikrlash va tezkor javob berish ko'nikmalarini shakllantiradi.



1-rasm. Wordwall saytidagi "Baraban" metodi.

Zamonaviy axborot-kommunikatsiya texnologiyalari asosida yaratilgan interaktiv o'quv metodlaridan yana biri sifatida "Kahoot" tizimi bugungi ta'lim jarayonida keng qo'llanilmoqda. Kahoot onlayn tarzda test, viktorina va interaktiv so'rovnomalar tuzish imkonini beruvchi bepul ta'limiy platformadir. Ushbu tizim orqali o'qituvchi real vaqt rejimida o'quvchilarga savollar beradi, o'quvchilar esa o'z qurilmalari (telefon, planshet, noutbuk) orqali javob beradilar. Natijalar darhol ekranda aks etadi va o'yin shaklida baholanadi. Kahoot platformasidan foydalanish uchun foydalanuvchi, avvalo, <https://kahoot.com> (<https://kahoot.com/>) veb-sayti orqali ro'yxatdan o'tishi lozim. Ro'yxatdan o'tish jarayonida foydalanuvchi o'zini "o'qituvchi" sifatida tanlaydi va shaxsiy ma'lumotlarini kiritgan holda akkaunt yaratadi. Tizimga Google, Yandex kabi qidiruv tizimlari orqali kirish imkoniyati mavjud. O'yin davomida savollar navbatma-navbat ekranda namoyish etiladi. Har bir savolga to'g'ri va tez javob bergan o'quvchilar reyting jadvalida yuqoriga ko'tariladi. Ushbu jarayon sog'lom raqobat muhitini yaratadi va o'quvchilarning motivatsiyasini oshiradi. Har bir bosqich yakunida natijalar avtomatik tarzda tahlil qilinadi.



2-rasm. "Kahoot" interaktiv o'yini.

Xulosa qilib aytganda, til ta'limida kompyuter texnologiyalaridan foydalanish keng imkoniyatlar yaratadi. Ushbu texnologiyalar o'quv jarayonining samaradorligini oshiradi, uni interaktiv va qiziqarli shaklda tashkil etishga xizmat qiladi. Raqamli vositalar orqali o'quvchilar bilimlarini mustahkamlash, xatolarini tezkorlik bilan aniqlash va bartaraf etish, shuningdek, individual yondashuv asosida ta'lim olish imkoniyatiga ega bo'ladilar. Bu esa, o'z navbatida, til o'rganish jarayonining yanada takomillashishiga xizmat qiladi. Shu bilan birga, zamonaviy axborot texnologiyalari o'qituvchilarga o'z pedagogik faoliyatlarini samarali tashkil etish uchun zarur vositalarni taqdim etadi. Ta'lim tizimida raqamli texnologiyalarni qo'llash o'quvchilarga individual yondashuv asosida ta'lim olish imkoniyatini yaratadi va ularning ijodiy, analitik fikrlash ko'nikmalarini rivojlantiradi. O'zbekistonning ta'lim tizimida zamonaviy texnologiyalarning joriy etilishi ta'limning global talablariga moslashish va ta'lim sifatini oshirishda muhim rol o'ynaydi.

Foydalanilgan adabiyotlar:

1. O'zbekiston Respublikasining "Ta'lim to'g'risida"gi Qonuni.
2. Muhamedova S.X., M. D. Abdullayeva, Sh. Sh. Yuldasheva, Y. B. Eshmatova "O'zbek tili va adabiyoti ta'limida zamonaviy axborot texnologiyalari" Toshkent, 2020.
3. Information and communication technologies in distance learning: Special training course / trans. from English Michael G. Moore, W. McIntosh, L. Black and others. M.: Publishing House "Education-Service", 2006.
4. Dmitrieva E.I. Didactic capabilities of computer telecommunication networks for teaching foreign languages // Foreign languages at school. 2007. -22 –26 b.
5. Нелунов А.И. Личностно-ориентированное обучение иностранному языку на основе информационных технологий // Информатика и образование. 2008.-31-33с.

6. Polipova T.A. Introduction of computer technologies in foreign language teaching // Foreign languages at school. 2009. -11-16 p.
7. Rahimov A.E. Chet tillarini o'qitishda kompyuter texnologiyalaridan foydalanishning afzalliklari // Maktabda chet tillari. 2012. -56–59 b.
8. Problems of application of multimedia technologies in higher education // High technology in pedagogical process: abstracts Interuniversity Scientific Conference of university teachers, researchers and specialists. Frolova N.H. Nizhni Novgorod, VSPI, 2000. -96 98b.
9. <https://www.google.com>
10. www.ziyonet.uz

THE ROLE OF CONFLICT IN THE STRUCTURE OF A LITERARY TEXT

Allashev Akbarbek Anvar ugli

Mamun university, Khiva, Uzbekistan

E-mail: akbar.allashev@gmail.com

ORCID ID: <https://orcid.org/0009-0007-6260-2436>

Abstract: The article highlights the distinctive relationships between a literary work's narrative, semantic core, and conflict, as well as the theoretical and literary idea of conflict and its function in a literary text. A thorough categorization of the many conflict types is provided. Such a phenomena of a literary text and its perception as conflict expectation are given special attention. It is noted that the storyline, the semantic kernel, and the conflict have an odd connection. A thorough categorization of the many conflict kinds is proposed. The literary work's phenomena like conflict anticipation receive special emphasis.

Keywords: conflict, plot, narrative character, oppositional composition lyayushaya, character, conflicting expectation.

He felt that tickling emptiness that always accompanied his desire to write.

In this emptiness, something took shape, grew.

Christmas, new, special. This old snow and new conflict...

Vladimir Nabokov, "A Christmas Story", 1928

Introduction

The most significant type of creative text is conflict, which should be seen as a collection of the author's unique thoughts about a certain number of circumstances and events that make up a story. Events and situations are different in that the former is a more dynamic aspect of the story, whilst the latter is more static. The way events and circumstances are arranged and alternated in a literary work is known as the plot, and it serves the function of establishing, expressing, and highlighting the text's semantic core. In this instance, a sequence, a collection of events, and situations alone cannot be interpreted as the plot; rather, they create a plot because of conflict (from the Latin *konfliktus* - clash). Only conflict gives a dynamic beginning to a series, a sequence of particular events and situations, creates the content of the artistic text.

Materials and methods

According to S. Sirotvinsky, conflict is "the engine of the development of action and counteraction in a dramatic or epic work" [1, 131–132]. He defines conflict as "the contact of opposing aspirations, the revelation of contradiction". Regarding the lyrical text, it is particularly important to remember the meaning of the term "text" (*textum*, which is translated from Latin as "connection").

Although it is unjustly left out in this instance, it should be highlighted that, given the particulars of the lyrical conflict, the conflict also plays a comparable function in it. A tradition has emerged in literary studies where conflict is understood to inevitably entail a clash, struggle, argument, expression of opposing opinions and polar inclinations, and a confrontation of warring forces, both in the heroes' inner world and in the outside world. The strongest indication that the world portrayed in a literary work is dualistic and that its dualism is oppositional is found in the very nature of conflict and the function it serves. The primary structural tenet of this resistance is the structure and existence of reality: spiritual and material, darkness and light, good and evil, earth and sky, friends and enemies...

However, we can say that conflict is not just a confrontation but also the nature of connections and a condition thanks to the contemporary interpretation of conflict in literary texts. In Lermontov's "I go out alone on the road..." for instance, there is no conflict at all if we disagree with this and view conflict just as a confrontation. However, the lyrical hero's realization of the essence of his situation, location in time, and space is reflected in the numeral "one" itself, which

is announced in the cited line. The poem "I loved you..." by A. S. Pushkin is a fitting example; it depicts not so much a conflict as the character of the connection between the lyrical hero and the addressee in the form in which they have developed in the present tense.

The "conflict-free nature of the narrative" in I. A. Bunin's novella "Antonov Apples" was once the subject of much writing. In the meanwhile, this narrative, like several others written during this time, is characterized by the desire to preserve what ought to be kept in the people's spiritual experience as well as to record the "connection of times". The ability to remember is therefore the primary indicator of the nature of the conflict, which is resolved positively, when statements like "I remember the early fine autumn", "I remember his old woman", "And I remember, sometimes it seemed to me" "I did not know or see serfdom, but I remember," etc., appear in the text. This is because the inability to remember is its exact opposite resolution.

In other words, a conflict presupposes the presence of two, most often oppositional, components, which in the perceiving consciousness can be associated: one with a positive, and the other with a negative emotional perception, which does not deny the possibility of the presence of positive potential in both components. For example, both lovers can represent a positive principle, which, by no means, excludes conflict situations that can arise between them. The role of such oppositional components can be performed by various structural elements of artistic text.

Result and discussion

Firstly, they can be two characters: a positive one (usually the main character) and his opponent and even his antipode (usually negative), when the main content of the conflict is connected with different worldviews, life positions and goals, different sense of self in time and space. This type of conflict components can be described in the words of A.S. Pushkin:

They came together: water and stone, poetry and prose, ice and fire... For example, the clash of two life positions is the basis of the plot of F. Schiller's tragedy "Mary Stuart" (1801). Henry Fielding in the novel "The History of Tom Jones, a Foundling" (1749) builds the narrative on the conflict between the poor but moral Tom, rejected by the father of his beloved, Squire Western, and the rich but immoral Blifil. Secondly, the components of the conflict can be the character and nature, when the hero enters into a struggle with natural, physical forces, most often exceeding his capabilities. Such is, for example, the conflict of W. Shakespeare's drama "The Tempest" (1612), Robinson Crusoe by Daniel Defoe and many of Jules Verne's hero-travelers are in such conflictual relations with nature. A conflict of this nature is the main basis for the development of plot of E. Hemingway's story "The Old Man and the Sea" (1952). A comic resolution of this conflict is presented in A. P. Chekhov's story "A Conversation between a Man and a Dog" (1885).

Thirdly, the role of such components can be played by two social ideological structures, taken at the moment of their collision, the goal of which is victory, the supremacy of one structure over the other. A family can act as such a social structure, like, for example, the Montagues and Capulets in W. Shakespeare's tragedy "Romeo and Juliet". The conflict of A. G. Malyshkin's story "The Fall of Dair" (1923) is built exclusively on the clash of two truly enormous human masses, reds and whites, two ideologies, when an individual, a character, practically dissolved in the mass. There is a white and a red army, there are regiments and "cavalry-partisan divisions", squadrons, there are army commanders, chiefs of staff, and almost no names of people are encountered. In the third part of D. S. Merezhkovsky's trilogy "Christ and Antichrist" (1895-1904), the novel "Antichrist. In Peter and Alexei (1904), the conflict outwardly appears as a clash between two heroes: Tsar Peter and Tsarevich Alexei, as a struggle between a persecutor and a guardian of the original Russian Orthodoxy, but in the general plan of the trilogy, this conflict is a clash of two worldviews, two truths, as Merezhkovsky himself expressively said: "When I began the trilogy Christ and Antichrist, it seemed to me that there were two truths: Christianity - the truth about heaven and paganism - the truth about earth, and in the future union of these two truths - the fullness of religious truth. But, finishing, I already knew that the union of Christ with the Antichrist is a blasphemous lie; I knew that both truths - about heaven and about earth - are already united in

Christ Jesus ... But now I also know that I had to go through this lie to the end in order to see the truth. From bifurcation to unification – this is my path – and my fellow reader, if he is equal to me in the main thing – in freedom of search – will come to the same truth” [2]. Fourthly, the opposing forces of the conflict can be natural phenomena (most often animals), the clash between which occurs without the direct participation of man. Such components of the conflict can be observed in the works of K. D. Ushinsky, Seton-Thompson, I. S. Turgenev, M. M. Prishvin. A conflict of such a structure can be conditionally called animalistic.

Fifthly, the opposing forces of the conflict can be presented as a character and society, when the conflict is built on the desire of an individual hero to realize his capabilities in the picture of the world proposed by the author and the difficulties, and most often the impossibility of such realization in a given time and space. This is how the relationships between Hamlet, King Lear and the world around them are built, such relationships with the surrounding reality are found in Stendhal's Julien Sorel and Flaubert's Emma Bovary, the hero of Balzac's novel Lost Illusions (1843) Lucien Chardon and Gorky's Foma Gordeyev... Sixth, such components of the artistic text can be the character and fate, when the conflict is built on the confrontation of man with the laws of fate or deity. Conflict with such components is usually called providential, it lies at the heart of the overwhelming majority of ancient tragedies. Seventh. Oppositional forces can be the qualities of the hero's inner world that come into conflict with each other, when the source of the conflict is the problem of the need to choose between duty and desire, conscience and needs, opportunity and necessity, etc. Such components of the conflict are most consistently presented in lyric poetry, but this does not mean at all that they are unknown to epic and drama. Such is, for example, the conflict in the story by I. V. Goethe “The Sorrows of Young Werther” (1774) and in the novel by F. M. Dostoevsky “Crime and Punishment” (1866); such is the nature of one of the main conflicts in the epic by M. A. Sholokhov “Quiet Flows the Don” (1940).

Eighthly. The components of the conflict can be the creator and creativity, when the artistic text becomes an image of what processes are taking place, what difficulties arise in the consciousness of the hero engaged in the creative process, what sorrows and joys the results of this process bring to the creator. Such a conflict became the basis, for example, of the lyrical reflections of V. A. Zhukovsky in “The Aeolian Harp”. Novalis's unfinished novel “Heinrich von Ofterdingen” (1802) is built on such a conflict. A significant part of the semantic complex of K. S. Aksakov's story “Walter Eisenberg” (1842) is also associated with an attempt to resolve a conflict of this nature. Literature knows numerous examples, including satirical ones, of the embodiment of this conflict, such as Sasha Cherny's poem “Overwork” (1908), which he dedicated to the “written-out “populists”:

I look like a woman in labor, I am ready to gnash my teeth... I curse the inkwell And the inkwell's mother! My hair is tousled, I have grown dull as a sheep, - Ah, all the rhymes have been spent To the end, to the end!.. I really have nothing to say today, as always. But believe me, I have never been embarrassed by this - I gave birth to little words and words, and gave birth to rhymes for them, And in cheerful verses, like a foal, I neighed [3, 89-90].

Ninth. The source of conflict in a literary text can be the contradiction between the spiritual and physical principles. One version of the resolution of this conflict is presented, for example, in Leo Tolstoy's story “Father Sergius” (1890-1898, published in 1912). In order to overcome his carnal desire for the woman who has come to his cell, the hero first thinks of that holy father “who laid one hand on a harlot and the other in a brazier,” but he does not have such a brazier. He puts his finger over the flame of the lamp, but cannot bear the physical pain for long. And then Father Sergius takes an axe, chops off his index finger, and then stands opposite the woman trying to seduce him and quietly asks: “What do you want?” [4, 360].

Another, directly opposite to the first, variant of resolving this conflict is also possible. It is presented, for example, in the story by A. P. Chekhov “Ionych”, when concern for the material, the carnal kills the spiritual principle in a person. The tenth type of conflict in a fictional text is built on a contradiction between the external manifestation of events and their essence, which is

implicit for the characters. It has a certain global character, because it is connected with the limitations of knowledge, experience, ideas of a person and even all of humanity, which are incapable of correctly understanding the essence of what is happening, of seeing the internal, deep essence in the external. An incorrectly understood essence as the internal content of ongoing processes can lead, on the one hand, to incorrect, erroneous decisions, and on the other, to fear of the surrounding world, an apathetic or painful perception of it. Thus, one of the heroes of world literature, Don Quixote, constantly makes wrong decisions and sometimes commits absolutely absurd actions, because the true meaning of phenomena and events is inaccessible to him, he cannot adequately interpret what is happening in real life. And the lyrical hero of Sergei Yesenin experiences mental anguish for the same reason: I am in a continuous smoke, In a life torn apart by a storm, That is why I suffer, because I do not understand Where the fate of events is taking us [5, 107-108]. It is noteworthy that one of the striking manifestations of this conflict for the artist is his own creativity, understood as a phenomenon of the surrounding life, the results and consequences of which cannot always be predicted and understood: "We are not given the ability to predict how our word will resonate...", F. I. Tyutchev once admitted [6, 199].

The first six types of conflict are usually defined as external, and the seventh, eighth and ninth are understood as internal. The tenth can be defined as synthetic, in which both external and internal components are combined and linked. In a literary text, conflicts of different types (internal, external, synthetic) can intersect, flow from one another, and be in a cause- and-effect relationship. The cause, the motivating principle of an internal conflict can be an external conflict and vice versa, but in any case, the entry of certain forces into conflict relations in a literary text is determined, conditioned by the presence of causes and goals. And the main determining principle is that the conflict of a literary text is the result of understanding the world as an oppositional structure. At the same time, it is the conflict that gives the literary text its main center of gravity, and therefore forms the author's semantic core. Therefore, conflict can be defined both as the main element of content and as the main means of embodying this content in a literary text. Thanks to conflict, its development and the results of this development, what is commonly called tension or weakening, acceleration or deceleration of action arising as a result of the clash of opposing forces arises in a literary text. Conflict in a literary text is the main source of formation, development, and disclosure of characters, which, in turn, also play the role of its source, "organizer," for this conflict. The nature of the conflict, which is the source and main reason for the development of the plot, determines the increase or decrease in the tension of the narrative, the presence or absence of elements that slow down its development (for example, descriptions, portraits or reasoning of the hero, author, narrator).

In the question of conflict and its role in the creation, development, and character of the plot, one can note one paradoxical, at first glance, feature. Despite the uniqueness of the plot of a work of art, this plot is built on the basis of conflict patterns that have been known to fiction for many centuries, they are historically repeated, constantly reproduced, each time receiving a new author- sian embodiment.

Since the plot of a work of fiction is built on well-known conflict patterns, the effect of conflict expectation can be observed in the relationship between the reader and the fiction text. It arises, first of all, in connection with the title of the work and is based on the reader's omniscience, on a certain foreknowledge, a prediction of how and on what the plot will be built. For example, if the reader is presented with a story called "Death on Capri", then even before starting to read, the reader already expects the resolution of the conflict that is stated in the title. The same Bunin story, but now with the title "The Gentleman from San Francisco" deprives the reader of his "omniscience", narrows the possibilities of foreknowledge. The reader who first turns to M. A. Sholokhov's epic "Quiet Flows the Don" and expects a story about "silence", and therefore "about peace and harmony in the space of the Don River", will be largely disappointed, his conflicting expectations will not be justified. Just as they will not be justified when reading Sholokhov's stories "Family Man" or "Native Blood" and others. Conflict expectation is the reader's anticipation of

upcoming events, which can either be destroyed as the reader reads, or confirmed. This conflict expectation or anticipation, foreknowledge is one of the foundations on which the entertainment of, of course, not only an epic work is built. And it applies not only to detective and adventure literature. Another thing is that for the latter, conflict expectation is of fundamental importance: there is little interest in reading a detective story when it is clear from the first pages how its conflict will end.

Conflict expectation can be stimulated by intertextuality declared in the title, when the reader's foreknowledge is based on the fact that he already knows the literary or historical-cultural phenomena that have become part of the title of the artistic text. For example, the story "Lady Macbeth of the Mtsensk District" by N. S. Leskov. S. Kozlov has stories "Russian Faust", "The Last Carthage", "Rehearsal of the Apocalypse".

Conflict expectation imparts a special interest to the artistic text: the reader is drawn into the new behavior of a well-known, old hero or into a new development of a well-known literary or historical conflict situation. However, the interest of an epic work cannot be understood only in a concrete-material sense, when the writer, and then the reader, are carried away by the acquisition of only some final knowledge about the resolution of the conflict. Interest, based on foreknowledge, anticipation, stems from the fact that it becomes interesting and fundamentally important how, from what side, foreknowledge is confirmed or refuted, how the old conflict scheme is technically constructed and resolved in a new, original, stylistically new way.

Conclusion

In conclusion, conflict expectation not only in the epic, but also in drama, and especially in lyric poetry creates a sense of harmony in the reader, a certain idea that the world of the work of art is not closed, not separated from him. What is happening in this world can be predicted, one can find in it both spiritual sympathy for the characters, and find one's own spiritual balance. And this same conflict expectation contributes to the formation in the reader's mind of the idea of the authenticity of the story. Based on the idea of opposition, contradiction between its constituent elements, the conflict determines the nature of the relationship between the characters and images of the work of art, as well as the stages of plot development. The conflict invariably puts not only the characters, but also the author in a position of choice, which can be philosophical, moral, logical, aesthetic. Both the author of a fiction text and its characters are forced to constantly choose between positions that constitute a conflict as an opposition, to choose, that is, to take the side of one of the positions: truth – lie, fair – unfair, logical – illogical, beautiful – ugly, decent – indecent, etc., etc. It is impossible in principle to avoid such a choice in a fiction text, due to the fact that it, as a conflict of the second reality, differs from the conflicts that a person encounters in the first reality. It differs, first of all, in that the conflict of a fiction text exists in time and space, which are characterized by discreteness and locality. Therefore, the choice can be postponed, transferred to another time and another space, the process of choice itself can be delayed, in principle, one can even refuse to choose a solution (and this is also a choice), but it is impossible to avoid the choice. Each of the positions that make up the opposition is understood in the text from the point of view of the category of evaluation, therefore the choice of solution turns out to be a justified and justified moment of self-determination – philosophical, ethical, logical, aesthetic.

References:

1. Sierotwinski S. Slownik terminow literackich. Wroclaw, 1966.
2. Merezhkovsky D. S. Introduction to the Complete Works. Published by the I. D. Sytin Partnership, 1914. Complete collected works. Vol. I. Moscow, 1914.
3. Black Sasha. Overwork // Black Sasha. Collected works: in 5 volumes. Vol. 1: Satires and lyrics. Poems 1905-1916. Moscow: Ellis Luck, 1996.
4. Tolstoy L. N. Father Sergius // Tolstoy L. N. Collected works. In 12 volumes. Vol. 10. Stories and short stories. 1872-1903. M., Art. lit., 1975.
5. Yesenin S. A. Letter to a Woman // Collected Works in 6 volumes. Vol. 2. Poems

- (1912-1925). Moscow, Khu- doj. lit., 1977.
6. Tyutchev F. I. We are not given to predict... // Works in 2 volumes. Vol. 1. Moscow, Pravda, 1980.
 7. Mohira, S. (2024). ISSUES, THEORIES, AND APPROACHES IN THE PRESENT-DAY LEXICOGRAPHIC SEMANTIC STUDIES. «CONTEMPORARY TECHNOLOGIES OF COMPUTATIONAL LINGUISTICS», 2(22.04), 438-442.

SEMANTIC CATEGORIZATION OF NOUNS IN THE UZBEK LANGUAGE THESAURUS DICTIONARY

Saparova Mohira

Senior teacher of Mamun university
Khiva, Uzbekistan

Abstract: This article describes the study of semantic fields, encompassing lexicographic principles, semantic-structural concepts, the origin and formation of semantic fields, their characteristic features, and the creation of semantic poles within the Uzbek thesaurus. Furthermore, it highlights that thesaurus, WordNet lexicons, and all forms of multilingual electronic dictionaries serve as the foundational database and core network for machine translation. The creation of the linguistic database and software for the Uzbek thesaurus involved an analysis of the semantic fields of lexemes and their conceptualization.

Keywords: semantic fields, lexicons, categorization, linguistic database, thesaurus, machine translation

Introduction

Thesaurus, WordNet lexicons, and all forms of multilingual electronic dictionaries serve as the foundational database and core network for machine translation. However, in the Uzbek language, these types of electronic dictionaries are not oriented towards natural language processing, and fully developed dictionaries of this kind have not been created in our linguistics. Academic articles and research have expressed considerations regarding the creation of UzWordNet and thesaurus lexicons for the Uzbek language. Our primary task is to compile the linguistic database for an Uzbek thesaurus and to develop its software. We have examined multilingual versions of this type of dictionary, and we are incorporating translations of words into English, Russian, German, and Tajik in Uzbek thesaurus. The structure of the Uzbek thesaurus is being developed as follows...”

- Developing/Building the linguistic database for the dictionary.
- Creating a software application based on a computer search engine. (Or, more concisely: Developing a search-engine-based software application).
- Identifying/Determining the semantic fields of nouns. (Or: Defining the semantic fields of the noun category).

Objectives.

Creating thesaurus is a task that demands considerable effort and the knowledge, skills, and expertise of specialists from various fields; it can be considered a massive laboratory undertaking.

We are enriching the linguistic base of the Uzbek thesaurus with lexemes belonging to the noun, adjective, and verb categories. A key component of language comprehension and the way we arrange our knowledge of the outside world is the semantic classification of nouns. We may learn more about the meaning of words and how they relate to one another by taking into account the many semantic categories. With applications in many other industries, it's a vast and intricate topic of research. For this research, we selected words belonging to the noun category as our object of study. Within the scope of this research, we have undertaken the following:

1. We extracted nouns from the explanatory dictionary of the Uzbek language and compiled them into a database. The results indicate that there are a total of 35,889 noun lexemes in the Uzbek language, of which approximately 1,000 have been analyzed to date.

2. We carried out a semantic categorization of the noun lexemes and noted the following:

The translation stops at point 2 because the description of the semantic categorization is missing from the provided text. To complete the translation, please provide the content of point 2. In this research, we attempted a general analysis of noun lexemes and categorized them as shown

in the table above: personal nouns, object nouns, place nouns, and activity/process nouns, which were then analyzed.

Scope and methodology

To align with the “Uzbek Thesaurus” dictionary's structure, we categorized each entry using the following fields: Word, Thesaurus Details, Categories, Example, Definition/Explanation, Group (Personal, Object, Place, or Activity/Process Noun), Synonym, Antonym, Hyponym, Hypernym, Meronym, Partonym, Holonym, Compound Word, Proverbs (minimum three), Proverbs-Related Phrases, Explanation of the Phrase, Idioms, Explanation of the Idiom, Idiom-Related Example, and Explanation of the Example. According to the requirements of the “Uzbek Thesaurus” dictionary, we grouped words into the following categories:

Four broader categories:

- I. Morphological level: word
- II. Semantic level: Synonym, antonym, hyponym, hypernym, meronym, patronym, holonym, compound word
- III. Syntactic level: Noun
- IV. Function: Examples, definition, explanation

• The final stage concludes with a “save” function.

The application area of noun lexemes was categorized as follows:

PERSONAL NOUNS:

- a) Education-related: teacher, specialist, trainee teacher, dean, pro-rector, department head
- b) Religious: mullah, qori (reciter of the Quran), preacher
- c) Profession-related: photographer, carpenter, tailor, cook, artist, jeweler, doctor
- d) Sports-related: coach, boxer, wrestler
- e) Commerce-related: seller, entrepreneur, businessman, broker
- f) Dynasty-related: Timurids, Astrakhanids, Abbasids
- g) Nationality-related: Uzbek, Kazakh, Russian, Tatar, Turkmen
- h) Judicial system-related: judge, lawyer, prosecutor, investigator, criminal, witness
- i) Kinship-related: aunt (maternal), aunt (paternal), nephew, uncle (maternal), uncle (paternal)
- j) Military-related: soldier, general, colonel, lieutenant colonel, commander
- k) Political governance-related: governor, deputy, mahalla elder, president
- l) Medical-related: doctor, nurse, urologist, obstetrician, cardiologist, therapist, ophthalmologist, neurologist, psychiatrist, dentist
- m) Lineage-related: eshon (respected elder), sheikh
- n) Social status-related: prince, princess

OBJECT NOUNS:

- a) School supplies: book, notebook, pen, pencil, ruler, eraser
- b) Kitchen utensils: knife, spoon, fork, bowl, cup, teapot, sieve, bucket
- c) Household items: wardrobe, carpet, lamp, pillow, seat, table, chair, window, door, bookshelf, curtain, sofa, lamp, curtain
- d) Clothing: shirt, glove, pants, hat, jacket, boots, sneakers, hat
- e) Vegetables: melon, watermelon, pumpkin, tomato, cabbage, cucumber
- f) Fruits: apple, peach, cherry, quince, pear, plum
- g) Dishes: pilaf, shashlik (kebab), samsa (pastry), manti (dumplings), sumalak (sweet rice pudding), qatlama (layered pastry), suzma (yogurt cheese), mastava (soup)
- h) Animals: tiger, wolf, fox, bear, rabbit, horse, cow, sheep, dog
- i) Poultry: chick, hen, rooster, turkey, goose, duck, stork
- j) Customs and traditions: good news, cradle ceremony, white dress ceremony, bride's greeting, yor-yor (traditional song), lullaby, wrestling, uloq (traditional game), darboz (traditional game), lazy wife, sumalak (sweet rice pudding); cradle celebration, circumcision, muchal (traditional ceremony), wedding, Navruz, holidays

PLACE NOUNS:

- a) Water bodies: irrigation ditch, pool, lake, puddle, river, sea
- b) Buildings: kindergarten, school, college, university, hospital, prison, shop, warehouse
- c) Fields: garden, field, apricot orchard, field of crops, orchard, lemon grove, flower garden
- d) Hills: hill, mountain, peak, plain, pasture, meadow

ACTIVITY/PROCESS NOUNS:

- a) Activity nouns: teaching, driving, animal husbandry, poultry farming, tailoring, medicine, carpentry
- b) Process nouns: harvest, election, mowing, yield, competition

This translation provides a comprehensive overview of the provided text. Remember that the accuracy of specific terms might depend on the precise nuances intended within the original Uzbek.

In English, the concept of deep structure is represented by the term “deep structure” and surface structure is represented by the term “surface structure”

The sentence “John hit the ball” is an example of a surface structure, while “John met Bill on the corner” could be considered an example of a deep structure. While the surface structure of a language only reflects grammatical agreement, the deep structure incorporates both grammar and a strong semantic component. For instance, in the sentence “The work is difficult”, the surface structure indicates the lexeme “is hot”, which is semantically inaccurate. Because semantic-structural analysis is a complex concept, this distinction plays a crucial role when translating from one language to another. For example, in the English sentence “The dog treed the cat”, the word “treed” is used as a verb, although it is fundamentally a noun lexeme. In such cases, semantic-structural analysis is helpful. The smallest unit in semantic-structural analysis is the semantic component. Semantic components combine to form concepts. The concept of meaning includes four principles:

Literature Review

The importance of the thesaurus as a research object in lexicography is emphasized. The article reviews the work of researchers who have explored specific aspects of thesaurus in lexicographical research, including: W. Teubert (on European corpus linguistics and its practical applications), H. Bergenholtz and H. Gouws (on discrepancies in defining the nature and scope of lexicography), and others who have provided theoretical and practical definitions of lexicography, with H. Bergenholtz and H. Gouws offering various interpretations. W. Klein’s work on lexical units and H. Jackson’s comprehensive study on the history of lexicography, its principles, critical approaches to lexicology, dictionary typology, structure, content, and usage are also mentioned. The importance of the thesaurus as a research object in lexicography is emphasized. Reflections on bilingual dictionaries, the study of phraseologisms in lexicography, and the semantic properties of lexemes are presented in T. Magay’s research. Henri Bejoint’s scholastic study compares British and American lexemes, pointing out issues with lexicography, word definitions, synonyms, and other meaning discrepancies. (Tagay, 2000;) K. Fellbaum mostly discusses the useful applications of corpus-based lexicographic techniques and digital lexicography using WordNet dictionaries. Among Uzbek academics, N. Abdurakhmonova has carried out several applied research in the fields of corpus-based translation studies, computational linguistics, and corpus-based language education, with a focus on creating lexicographic databases. Using an English-Uzbek machine translation module as an example, she has also shed light on problems pertaining to translation technology, algorithms, and the development of linguistic resources.

Result and Discussion

1. Object
2. Event
3. Attribute (Specificity)
4. Relationship

In the Uzbek thesaurus we are creating, the semantic-structural analysis and semantic fields are integrated, with words initially categorized by field, group, and category during their creation. Focusing primarily on the noun category, semantic fields are generated as follows: For instance, when analyzing abstract nouns, the semantic field initially encompasses synonymous concepts.

Let's consider the semantic fields of the word "baxt" (happiness), which is included in our Uzbek thesaurus.

Here, the lexeme "baxt" is combined in the semantic fields of tole (fortune), saodat (bliss), qut (blessing), mamnuniyat (satisfaction), rohat (pleasure).

It is worth noting that Sh. Safarov and M. Mirtojiev have thoroughly studied Uzbek semantics and semasiology and conducted numerous studies in this area. The central issue in semantics is noting phenomena related to lexical meaning and its development.

Findings

In our Uzbek thesaurus, special emphasis is given to noun lexemes expressing activity and process, which form semantic fields as follows:

The text cuts off before listing the semantic fields for activity and process nouns. To complete the translation, please provide the missing information. In addition, with the aim of further enriching Uzbek lexicography, we have created an Uzbek thesaurus based on the Princeton WordNet and are continuing to expand it with words. Our main goal is to demonstrate the richness of our language and the multifaceted meanings of each word. Unlike the explanatory dictionary of the Uzbek language, the Uzbek thesaurus includes the English and Russian translations of each word, its definition, examples, idioms and proverbs with their explanations. Furthermore, it provides hierarchical concepts such as antonyms, synonyms, partonyms, meronyms, holonyms, hyponyms, and hypernyms for each word. As a result, using the English WordNet thesaurus as a model, a database was created on the website <http://thesaurus.uzbekcorpus.uz/uz>, focusing on noun lexemes, exemplified by the concept of "family values". The lexemes primarily associated with the family concept were analyzed.

Limitations and Research gaps

Linguocultural markers play a significant role in determining the semantic characteristics of nouns. The inclusion of linguistically and culturally specific words in the Uzbek thesaurus presented several challenges. For example, with the noun "somsa" (a type of pastry), which falls under the concept of food, creating hyponyms or including other nouns related to clothing (like "doppi" - hat, "tandir" - tandoor oven, "zarchapon" - type of traditional garment) presents difficulties because direct equivalents may not exist in other languages, nor are these concepts always represented in WordNet. This highlights the need for careful consideration when developing the linguistic base of the Uzbek thesaurus. Challenges in translating linguoculturally specific words include finding linguistic equivalents, creating alternatives within the target language, and preserving the cultural specificity of terms.

Conclusion

Thesaurus dictionaries are valuable resources for information retrieval, significantly facilitating the indexing of lexemes and the identification of appropriate words. Thesauri are essential for selecting the most suitable terms during both indexing and word searching, highlighting the practical importance of automatic thesaurus. Semantic and grammatical analysis of language models necessitates the creation of computer-compatible electronic databases that are continuously refined using corpora. WordNet, analyzed and researched extensively by global scholars using diverse approaches, aids in understanding the nature of thesaurus. WordNet is a database containing semantic and lexical relationships between words. The basic unit in WordNet is not a single word but a synset—a set of synonyms forming a semantic network node representing similar meanings. WordNet is a database containing the lexical resources of the English language. In this database, nouns, adjectives, verbs, and other words are grouped into sets of cognitive synonyms (synsets) and individually defined. WordNet is publicly available for download and serves as a valuable resource for creating a thesaurus in Uzbek. Thesaurus illuminate semantic

relationships between words, including semantic relationships, form-meaning relationships, usage in phrases, near-synonymous words, hypernyms, hyponyms, meronyms, and holonyms. Such dictionaries allow for a broader representation of the capabilities of the Uzbek language. While creating a hierarchical taxonomy of all lexical resources within a specific domain in a thesaurus, a network of the entire language lexicon is necessary in linguistic ontology. Therefore, creating thesauri is relatively more efficient and faster than creating ontologies. Creating a word network from scratch or expanding an existing one is a laborious process involving multiple stages and extensive use of both human effort and automated systems. The creation of the Uzbek thesaurus has enriched the Uzbek language and expanded its capabilities. Challenges encountered in incorporating hierarchical taxonomies into the dictionary were resolved through human intervention.

References:

1. Abduraxmonova N.Z. O‘zbek tili elektron korpusining kompyuter modellari. Filologiya fanlari doktori (DSc). – T., 2021. – 88 b.
2. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, 11(19), 9117.
3. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
4. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
5. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: [uzbekcorpus. uz](http://uzbekcorpus.uz). In *2021 International Conference on Information Science and Communications Technologies (ICISCT)* (pp. 1-4). IEEE.
6. Bates M. (1995). “Models of natural language understanding”. *Proceedings of the National Academy of Sciences of the United States of America*. 92 (22): 9977–9982. Bibcode: 1995PNAS... 92.9977B. doi:10.1073/pnas.92.22.9977. PMC 40721. PMID 7479812.
7. Alessandro Agostini, Timur Usmanov, Ulugbek Khamdamov, Nilufar Abdurakhmonova, and Mukhammadsaid Mamasaidov. 2021. UZWORDNET: A Lexical-Semantic Database for the Uzbek Language. In *Proceedings of the 11th Global WordNet Conference*, pages 8–19, University of South Africa (UNISA). Global WordNet Association.
8. Bates M. (1995). “Models of natural language understanding”. *Proceedings of the National Academy of Sciences of the United States of America*. 92 (22): 9977–9982. Bibcode: 1995PNAS... 92.9977B. doi:10.1073/pnas.92.22.9977. PMC 40721. PMID 7479812.
9. Alessandro Agostini, Timur Usmanov, Ulugbek Khamdamov, Nilufar Abdurakhmonova, and Mukhammadsaid Mamasaidov. 2021. UZWORDNET: A Lexical-Semantic Database for the Uzbek Language. In *Proceedings of the 11th Global WordNet Conference*, pages 8–19, University of South Africa (UNISA). Global WordNet Association.
10. Bates M. (1995). “Models of natural language understanding”. *Proceedings of the National Academy of Sciences of the United States of America*. 92 (22): 9977–9982. Bibcode: 1995PNAS... 92.9977B. doi:10.1073/pnas.92.22.9977. PMC 40721. PMID 7479812.

BADIIY MATN TAHLILIDA OKKAZIONALIZMLARNING LINGVOPOETIK IMKONIYATLARI

Inogamova Nargiza Shavkatovna

O'zbekiston Milliy universiteti
Kompyuter lingvistikasi va amaliy tilshunoslik
kafedrası katta o'qituvchisi

Annotatsiya: Lingvopoetika til birliklarining badiiy asarlardagi estetik vazifasi va poetik imkoniyatlarini o'rganadi. Bu tahlil usuli matnning chuqur ma'no qatlamlarini ochishga, muallifning uslubi va ijodiy mahoratini anglashga xizmat qiladi. Maqolada badiiy matn tahlilida okkazonalizmlarning lingvopoetik imkoniyatlari xususida so'z boradi. Okkazonalizmlarning badiiy asar emotsional-ekspressivligini ta'minlashdagi o'rni, ularning yasaliishi xususidagi fikrlar bayon qilinadi. Badiiy matn tahlilida okkazonalizmlarning lingvopoetik imkoniyatlari O'zbekiston xalq shoiri Iqbol Mirzo she'riyati misolida tahlilga tortilgan.

Kalit so'zlar: badiiy matn, badiiy matn tahlili, okkazonalizm, lingvopoetika, lingvopoetik imkoniyatlar, okkazonal yasalmalar, matnning emotsional-ekspressivligi, hissiy bo'yoqdorlik, matndagi tushunchalarni obrazli nomlash.

Аннотация: Изучение языка художественных произведений является центральной проблемой лингвопоетики. В них разрабатываются различные методы и способы анализа художественного текста, среди которых можно выделить два основных подхода: лингвистический и литературоведческий. В данной статье излагаются лингвопоэтические возможности окказионализмов. Изучается эстетическая функция окказионализмов и их словообразование. Лингвопоэтические возможности окказионализмов анализированы на примере стихотворений народного поэта Узбекистана Икбола Мирзо.

Ключевые слова: художественный текст, анализ художественного текста, лингвопоэтика, окказионализмы, лингвопоэтическая значимость, эстетическую функцию языковых единиц, эстетические свойства.

Annotation: Linguopoetics studies the aesthetic function and poetic potential of linguistic units in literary works. This method of analysis helps uncover the deep layers of meaning within a text and serves to understand the author's style and creative mastery. The article discusses the linguopoetic potential of occasionalisms in the analysis of literary texts. It presents views on the role of occasionalisms in ensuring the emotional-expressive quality of a literary work and their formation. The linguopoetic potential of occasionalisms in literary text analysis is examined through the example of the poetry of People's Poet of Uzbekistan, Iqbol Mirzo.

Keywords: literary text, literary text analysis, occasionalism, linguopoetics, linguopoetic potential, occasional formations, emotional-expressiveness of the text, emotional coloring, figurative naming of concepts in the text.

Keyingi yillarda badiiy asar tilini lingvopoetik jihatdan tahlilga tortish ko'pgina ilmiy tadqiqotlar markazida bo'lib kelmoqda. Negaki badiiy matnni tahlil qilishda lingvopoetik yondashuv muhim ahamiyatga ega. Ma'lumki, lingvopoetika til birliklarining badiiy asarlardagi estetik vazifasi va poetik imkoniyatlarini o'rganadi. Bu tahlil usuli matnning chuqur ma'no qatlamlarini ochishga, muallifning uslubi va ijodiy mahoratini anglashga xizmat qiladi. Shu bois bunday tadqiqotlar asosida tilshunoslarimiz tomonidan badiiy matnni til va adabiyot nuqtayi nazaridan tadqiq etishda qo'llaniladigan usullar ishlab chiqilmoqda. Xususan, badiiy matn lingvistik tadqiqida so'zning estetik ta'siri, ma'no va mazmun uyg'unligi, lingvopoetik tahlil olib borishda muallifning individual uslubi, badiiy matnning mohiyati va estetik vazifasi, badiiy matn

estetik ta'sirini oshiruvchi badiiy aktuallashtirgan leksik vositalarning tahlili kabi masalalarga alohida to'xtalish lozim.

Dunyo tilshunosligida ham o'zbek tilshunosligida ham shu kunga qadar lingvopetika sohasida qator tadqiqotlar amalga oshirilgan, monografiyalar chop etilgan. Lingvopoetikaga oid ishlarda ijodkorning mahoratini ko'rsatuvchi vositalar ko'chimlar, metaforalar qatorida okkazional yasalmalarga ham o'rin berilgan ishlar mavjud. Jahon tilshunosligida B.A.Belova, V.V.Lopatin, E.A.Zemskaya, N.I.Feldman⁹³ tadqiqotlarida okkazionalizmlarga oid qimmatli fikrlar berilgan. Badiiy matn tahlili, so'z yasaliishi tadqiqiga oid ishlarda ham okkazional yasalmalarga tadqiqotchilar to'xtalib o'tganlar. O'zbek tilshunosligida S.Toshaliyeva⁹⁴ birinchilardan bo'lib okkazionalizmlarning yasaliishiga doir tadqiqotini amalga oshirgan. O.To'xtasinovanning "O'zbek tilida leksik okkazionalizmlar va ularning badiiy-estetik xususiyatlari" mavzusidagi tadqiqoti okkazionalizmlarning xususiyatlari, yaratilishi masalalariga qaratilgan. [To'xtasinova O. 2007 – B.118]

A.Hojiyevning "Tilshunoslik terminlarning izohli lug'ati"da okkazionalizmlarga quyidagicha ta'rifi beriladi: "Okkazionalizmlar. Mahsuldor bo'lmagan model asosida yasalgan va faqat shu nutqiy matnning o'zida qo'llangan so'z, individual-uslubiy neologizmlardir"⁹⁵. [Хожиев А. 1985. – Б.74.]

Neologizmlar endi paydo bo'lganida yangilik bo'yog'ini saqlab turadi, til iste'molchilari tomonidan o'zlashtirilganidan so'ng yangilik bo'yog'ini yo'qotib, me'yoriy so'zlarga qatoriga o'tadi. Faqat shu soha bilan shug'ullanuvchilar va tilshunoslarga bu so'zlarni qachonlardir yangi so'z bo'lganligini eslab qoladilar. Okkazionalizmlar esa neologizmlarga nisbatan badiiy matnlarda yashaydi. Okkazionalizmlar ham neologizmlar singari badiiy asarlarda uslubiy vazifani bajaradi, ammo ularning badiiy-estetik vazifasi yuqoriroq bo'ladi. Okkazional yasalmalar nutq jarayonida muhim uslubiy vazifa bajarganidek, tilning leksik tarkibini ham boyitib boradi. Badiiy asarlar, ayniqsa, she'riyat badiiy okkazionalizmlar yaratiladigan maydon hisoblanadi. Shoir yoki yozuvchi o'zining xayoliy dunyosini kitobxonga yetkazish uchun izlanadi. Okkazionalizmlarning asosiy vazifasi aniq badiiy nutqiy holat uchun xizmat qilish bo'lib, ular til faktiga aylanishi yoki aylanmasligi mumkin. Ularning qimmati bu bilan emas, kontekstda o'zining badiiy funksiyasini qanday bajarganligi bilan o'lchanadi. Badiiy adabiyotda nafaqat mazmun, balki so'zning qanday jaranglashi ham muhim rol o'ynaydi. Mazmun va jarangdorlik birgalikda ta'sirchanlikni vujudga keltiradi. Demak, badiiy nutqda asosiy e'tibor ana shularga qaratiladi. Badiiy nutqda ishlatilgan okkazionalizmlarni o'rganishdan shu narsa ma'lum bo'ladiki, ular ma'lum maqsadlarda, xususan, nutqning mazmunini barcha qirralari bilan aks ettirish, ifodalanayotgan voqea-hodisani obrazli tasvirlash, muallif bermoqchi bo'lgan fikr va g'oyalarni aniq ifodalash, matnning emotsional-ekspressivligi, hissiy bo'yodqorligini oshirish, berilayotgan matndagi biror voqea-hodisa, mavhum tushunchalar, joy va shaxslarni obrazli nomlash maqsadida qo'llanadi.

Maqolamizda O'zbekiston xalq shoiri Iqbol Mirzo she'riyati misolida so'zning badiiy asar qimmatini oshirishdagi muhim jihatlar ko'rib chiqish maqsad qilindi, xususan, shoir asarlarida qo'llangan okkazionalizmlar, okkazional yasalmalar tahlilga tortildi.

Iqbol Mirzo o'ziga xos so'z qo'llash mahoratiga ega bo'lgan shoirlardan biri. Uning she'rlari xalqona ohanglarda bitilganligi bilan boshqa shoirlarning ijod namunalaridan ajralib turadi. Bundan bir necha yillar avval dunyo tan olgan taniqli yozuvchi Chingiz Aytmatov Iqbol Mirzo timsolida zamonaviy o'zbek she'riyati iqbolini ko'rayotganligini aytgan edi. Yillar o'tib buyuk adibning fikrlari o'z tasdig'ini topmoqda. Iqbol Mirzoning qator she'rlari xalq orasida

⁹³ Белова Б.А. Об особенностях компонентного состава индивидуально-авторских сложных прилагательных (на материале поэзии В.Луговского) // Русское слово и языке и речи. — Кемерово: Кемеровского ГУ, 1976. — с. 3; Лопатин В.В. Рождение слова // Неологизмы и окказиональное словообразование. — М.: Наука, 1973. — с. 67; Земская Е.А. Современный русский язык. Словообразование — М.: Просвещение, 1973. — с. 229; Фельдман Н.И. Окказиональные слова и лексикография // Вопросы языкознания-1957-Н. С.66.

⁹⁴ Тошалиева С. Ўзбек тилида окказионал сўз ясалиши: Филол. фанлари номзоди ... дис. — Тошкент, 1998. — Б.166.

⁹⁵ Хожиев А. Тилшунослик терминларининг изоҳли луғати. ЎЗМЭ Давлат илмий нашриёти: — Тошкент, 1985. — Б.74.

xalqchilligi, o'ziga xosligi bilan mashhur bo'lib, o'z muxlislariga ega bo'lib kelmoqda. Uning she'rlarida poetik nutqining serqirraliligini ko'rsatuvchi vositalardan biri bo'lgan okkazional yasalmalar talaygina. **Qadimshunoslar, odamshunoslar, toshyigit, toshqiz, tilsimto'sh, yomg'irzor, bog'iston, nozbo'y** kabi yasalmalar shular jumlasidandir.

Okkazionalizmlarning yasalishiga e'tiborni qaratamiz. Shoirning "Yaproq" nomli she'rida **qadimshunoslar, odamshunoslar, toshyigit, toshqiz** kabi okkazionalizmlar uchraydi. – **shunos** biror soha bilan shug'ullanuvchi shaxs oti yasovchi qo'shimcha. **Qadimshunos** yasalmasi **tarixshunos** leksemasiga ma'nodosh. Lekin shoir she'rda keltirilgan voqeaning o'tmishga taalluqli ekaniga e'tiborni qaratib, arxeologik tadqiqotlar olib boruvchi shaxslarga nisbatan qadimshunoslar, odamshunoslar yasalmasini o'ylab topadi.

Okkazional so'zlarning yasali'sh tizimining bir necha turlari mavjud: muayyan so'z yasali'sh andozalariga muvofiq yasali'sh: *qadimshunoslar, odamshunoslar, yomg'irzor, xarobazor, bog'iston*.

Bu yasali'sh tizimida faol qo'shimchalar *-zor, -kor, -goh, -iston, -dosh, -kash, -shunos* kabi qo'shimchalardir. Bu qo'shimchalar asosga qo'shilib o'z vazifasini bajarsada, odatdagidan farq qiladi. Masalan, *-zor* qo'shimchasi til mezonlari bo'yicha jonli, o'suvchi narsalarni bildiradigan so'zlarga qo'shiladi, asosda anglashilgan narsa ko'p bo'ladigan joy nomini bildiradigan o'rin-joy otini yasaydi, lekin okkazional so'zlar tarkibidagi *-zor* qo'shimchasi ba'zan jonli, o'suvchi narsalarni bildiradigan so'zlarga emas, asosan mavhum otlarga qo'shilgan va bu tasodifiylik okkazoanal so'zlarni vujudga keltirgan:

Uch kunda kuz keldi. Dunyo qaridi.

Teraklar bosh chayqab qoldilar ma'yus.

Toshlar o'rmalaydi anhor qa'rida,

Anhor qaylarga'dir shoshadi hanuz.

Ko'pirib boradi qo'ng'ir bulutlar,

Hademay yomg'irzor bo'ladi paydo.

Yerga quloq bosib yotadi o'tlar,

Yerdanmi keladi bu mungli navo? (Kuz)

Tabiat bir mo'jiza. Inson shu tabiatning bir bo'lagi. Fasllar almashinuvi, bu jarayonda tabiatda bo'ladigan o'zgarishlar insoniyatni hamisha lol qoldirib kelgan. Adiblarimiz ijodida fasllar almashinuvini inson umrining ma'lum davrlariga qiyoslanish hollari ko'p uchraydi. She'rda uch kunda kuz kelishi umrning qisqaligiga, dunyoning qarishi yoshlikning abadiy emasligiga, anhorning shoshishi shiddat bilan o'tib borayotgan hayotda insonning ezgu ishlarni qilib ulgurishiga ishora qilayotgandek tuyuladi. Ko'pirib borayotgan bulutlardan yomg'irzorning paydo bo'lishi hademay boshlanadigan yog'ingarchilikka ishora. Misralardagi **yomg'irzor** so'zi okkazionalizm bo'lib, she'rning ta'sir kuchini yanada oshirgan. Aynan shu okkazional yasalma she'rdagi yashirin ma'noni aniqlashga yordam bergan. Shoir ifodalamoqchi bo'lgan kuz fasli bilan bog'liq tabiatdagi o'zgarishlar **yomg'irzor** okkazionalizmi orqali inson tasavvuridagi mavjud tabiat o'zgarishlarini umumlashtirib, yanada chuqurroq his etishga undagan. Yomg'ir aslida ko'z yoshiga ishora. Yomg'irzor okkazionalizmida o'tayotgan umrni to'xtatib qola olmaslikdan, chorasizlikdan oqayotgan ko'z yoshlarga ishorani angllamoq kerak.

Muqaddas yering bor, qutlug'osmon bor,

Suyangan tog'ing bor, dilda iymon bor.

Ulug' bog'istonda buyuk bog'bon bor,

Kaftingga ko'z suray, dehqonim, yurtim,

O'zingdan o'rgulay, bog'bonim, yurtim. (Navro'z qo'shig'i)

Keltirilgan parchadagi "**bog'iston**" yasalmasi ham Iqbol Mirzoning o'ziga xos okkazoional ma'nodagi topilmasi bo'lib, she'rning badiiyati va ta'sirchanligini oshirishga xizmat qilgan. Okkazoional ma'no – leksemaning tildagi ma'nosiga xos bo'lmagan ayrim shaxsning (muallifning) nutqiy vaziyatidan kelib chiqib, shu leksema mazmuniga yangicha tus berishi natijasida yuzaga keltirilgan sun'iy ma'no.

Bog' so'ziga –iston joy nomi yasovchi qo'shimcha qo'shilishi natijasida *bog'iston* leksemasi hosil bo'lgan. Bu leksemaning qo'llanilishida metafora usulida ma'no ko'chishiga guvoh bo'lamiz, ya'ni shoir tug'ilib o'sgan yurtini *bog'lar jamlangan makonga*, ya'ni *bog'istonga* o'xshatadi. Garchi bu yasalma xalq orasida ommalashib ketmagan bo'lsada, uning shu matndagi ta'siri sezilarli darajada yuqori ekanligi e'tiborni tortadi. Bu kabi topilmalar Iqbol Mirzo she'rlarida ko'plab uchraydi va ular o'ziga xosligi bilan ajralib turadi.

Okkazonalizmlarning keyingi yasaliş turi so'z qo'shish yoki so'z birikmasini lug'aviy birlikka aylantirish yo'li bilan yasaliş. Shoir ijodida uchraydidan *bozoryo'li, toshyigit, toshqiz, tilsimtosh* kabi yasalmalar so'z birikmasini lug'aviy birlikka aylantirish yo'li bilan yasalgan.

Keyin tushib ketding, sezmadimgo'yo.

Qayrilib boqmading, ey, sun'iy yo'ldosh.

Meni ayadingmi, o'zingnimi yo?

Qiz qalbi qizg'aldoq emas, tilsimtosh! (Metro)

Alanglaysan oyoq uchida,

Qiyofalar begona bari,

Bozoryo'li – qaynoq ko'chadan

Meni topolmaysan, dilbarim. (Alanglaysan oyoq uchida)

Yuqorida keltirilgan misollardagi, *tilsimtosh* okkazonal so'zi kompozitsiya usuli orqali yasalgan bo'lsa, "Alanglaysan oyoq uchida" she'ridagi *bozoryo'li* yasalmasi birikmalarini lug'aviy birlikka aylantirish yo'li bilan yasalgan va bu g'ayriodatiy so'zlar badiiy tasvirni yanada oshirgan. Xususan, *bozoryo'li* okkazonal so'zi bozorga olib boruvchi yo'l, bozorning yo'li va bu yo'lning qaynoqligi, ya'ni gavjumligini badiiy jihatdan ochib berishga xizmat qilgan. Bunday okkazonal so'zlar ijodkorning fikrini va o'ylarini o'quvchiga yanada aniqroq, ta'sirli va kengroq tarzda yetkazib berishga xizmat qiladi.

Rangi o'chdi atirgullarning, Xayr....

Chamanimdan nozbo'ylar ketdi. Xayr...

Oq qoraga ko'chdi kunlarim, Xayr...

Achchiq sog'inchi tunlari yetdi...Xayr... (Xayr)

Iqbol Mirzo ijodida *atirgul* obraziga murojaatlar talaygina. Atirgul visol, sevgi-muhabbat ramzi. She'rda atirgul rangining o'chishi, oshiqning chamanidan *nozbo'ylarning* ketishi, hayotdan mazmun, shodlik va quvonchning ketishiga ishora. Shoir metaforalardan mohirona foydalanib, oshiq qalbidagi tug'onlarni, his-tuyg'ularni obrazlar orqali mahorat bilan yetkazadi. Xijron, ayriliq motivlari bilan yo'g'rilgan ushbu misralardagi *nozbo'ylar* yasalmasi ham birikmalarni lug'aviy birlikka aylantirish yo'li bilan yasalgan bo'lib, badiiy matnning emotsional-ekspressiv bo'yog'ini oshirgan. *Chamanimdan nozbo'ylar ketdi.* Chamanzor gulzor so'ziga ma'nodosh. Gulzor leksemasi inson tasavvurida xushbo'y hid taraluvchi joy nomi sifatida gavdalanadi. Xushbo'ylik insonni sarmast qiladi. Salbiy tuyg'ularning ortga chekinib, xushkayfiyatlik holatiga olib kiradi. *Bo'y* leksemasi ham *Hid* leksemasiga ma'nodosh bo'lgani bois shoir qahramonning ruhiy holatidagi emotsionallikni oshirish maqsadida *bo'y* leksemasidan foydalanadi. *Noz* leksemasining mazmun planida yoqimli, nozik kabi semalar yotadi. Bu ikki leksemaning birlashuvi asosida *nozbo'ylar* okkazonal yasalmasi hosil bo'ladi va bu yasalma badiiy matnning mazmun planining ekspressivligini oshiradi.

Xulosa o'rinda aytish mumkinki, badiiy matnning lingvopetik tahlili shu tilning poetik imkoniyatlarini ochishga, muallif tomonidan til xazinasidagi mavjud til boyliklaridan foydalanish mahoratiga baho berishga xizmat qiladi. Iqbol Mirzo she'riyatini lingvopoetik jihatdan tahlil qilinishi, xususan, uning okkazonalizmlardan foydalanish mahorati o'ziga xos ekanligi shoirning original uslubni yarata olganligidan dalolatdir. Okkazonalizmlar badiiy asar tilini boyitishda, ifodaning obrazlilikini ta'minlashda, ta'sirchanligini oshirishda muhim ahamiyat kasb etishi shoirning ijod namunalarida o'z aksini topgan. Iqbol Mirzo she'rlarining o'quvchi qalbidan chuqur o'rin olishi, uning ruhiyatiga, hissiyotiga kuchli ta'sir qilishining sababi ham mana shunda. Iqbol Mirzo she'riyatining lingvopoetik tahlil qilinishi o'zbek adabiy tilining, jumladan,

o‘zbek poetik nutqining o‘ziga xos jihatlarini, betakror imkoniyatlarini ochib beradi. Bu esa o‘z o‘rnida o‘zbek tilidagi badiiy matnlarning betakror imkoniyatlarining ochib berilishiga imkon yaratadi.

Foydalanilgan adabiyotlar:

1. Iqbol Mirzo. Ko‘zlaringga termulsaydim. “Sharq” nashriyoti. Toshkent: 2020 –B.257.
2. Iqbol Mirzo. Ko‘nglingda nima bor, bilmayman. “Sharq nashriyoti”. Toshkent. 2020 – B.252.
3. Тошалиева С. Ўзбек тилида окказионал сўз ясалиши: Филол. фанлари ном...дис. – Тошкент, 1998. –Б.166.
4. Тўхтасинова О. Ўзбек тилида лексик окказионализмлар ва уларнинг бадий-эстетик хусусиятлари: Филол. фан. бўйича фалс. д-ри (PhD) дисс. – Тошкент, 2007 – Б.118.
5. Ҳожиев А. Тилшунослик терминларининг изоҳли луғати. ЎЗМЭ Давлат илмий нашриёти: — Тошкент, 1985. – Б.74.
6. Qurbonova M., Yo‘ldoshev M. Matn tilshunosligi. Toshkent: Universitet, 2014.

TARJIMA MATNLARI VA ULARDA MILLIY QADRIYATLARNING TAHLILI (Abdulla Qodiriyning “O‘tkan kunlar” romani misolida)

Abduraxmonova Sohiba Aktam qizi

Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti,
Jurnalistika va o‘zbek filologiyasi fakulteti, o‘qituvchisi
E-mail: abduraxmonovasohiba@gmail.com

Yashiboyeva Nozima Sirojiddin qizi

Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy Universiteti,
Jurnalistika va o‘zbek filologiyasi fakulteti,
Filologiya va tillarni o‘qitish(o‘zbek tili) yo‘nalishi 4-kurs talabasi
E-mail: nozimayaxshiboyeva3@gmail.com

Annotatsiya: Ushbu maqolada tarjima matnlar bilan ishlash va badiiy matnlarni tarjima qilish jarayonida milliy qadriyatlarning o‘rni haqida ma’lumot berilgan. Shuningdek, Abdulla Qodiriyning “O‘tkan kunlar” romanining amerikalik tarjimon Mark Reese tomonidan tarjima qilingan nusxasidagi milliy qadriyatlar bilan bog‘liq misollar tahlil qilingan.

Kalit so‘zlar: Tarjimachilik, milliy qadriyatlar, madaniyat, lingvokulturologiya, tarjimon.

Аннотация: В статье представлена информация о роли национальных ценностей в работе с переводными текстами и в процессе перевода художественных текстов. Также были проанализированы примеры, связанные с национальными ценностями в переводе романа Абдуллы Кадыри “Дни минувших дней” американским переводчиком Марком Ризом.

Ключевые слова: Перевод, национальные ценности, культура, лингвокультурология, переводчик.

Annotation: This article provides information about the role of national values in working with translated texts and the process of translating literary texts. Also, examples related to national values in the translation of Abdulla Qodiri’s novel “Days Past” by American translator Mark Reese were analyzed

Keywords: Translation, national values, culture, linguoculturology, translator.

Tarjima o‘zining ko‘p qirraliligi bilan dunyoni to‘g‘ri anglash va madaniyatlararo o‘zaro muloqotda muhim ahamiyat kasb etadi. Ammo turli mamlakatlardan kelgan mutaxassislar o‘rtasida hali hanuz til to‘siqlari va tarjima jarayonidagi qiyinchiliklar mavjud. Shu bilan bir qatorda, mashina tarjimasi jarayonida ham ayrim kamchiliklar ko‘zga tashlanayotgani sir emas. Chunki bir tildagi badiiy matnni boshqa tilga o‘girish jarayonida madaniyatlar orasidagi farq tarjima jarayonida bir muncha qiyinchiliklar keltirib chiqarishi mumkin. Masalan, o‘zbek tilidagi matnlarni tahrir qilish bugungi kungacha inson mehnati bilan bajarib kelinmoqda. Bu matnlarni tahrir qilishga ketadigan vaqtning ortishiga, iqtisodiy xarajatlarning ko‘payishiga, inson tomonidan qilinadigan ayrim xatoliklarga olib keladi.

Tarjimaning asosiy maqsadi va xususiyatlari turli mamlakatlar xalqlar o‘rtasidagi o‘zaro tushunishni rivojlantirishdan iborat bo‘lib, u madaniyatlarni bog‘lovchi ko‘prik vazifasini bajaradi. Shu o‘rinda mashhur amerikalik tarjima nazariyachisi Yevgeniy Nida tarjimaga quyidagicha ta’rif berganligini ko‘ramiz “Tarjima, birinchidan, ma’no jihatdan, ikkinchidan, uslub jihatdan qabul qiluvchi tilni manba tilning sof ekvivalenti sifatida takrorlashdan iborat. Biroq tarjima qilinayotgan tilda eng yaqin o‘xshash ekvivalenti hosil bo‘lishida madaniy farqlar turli

darajada ta'sir ko'rsatadi"⁹⁶. Turli xalqlar bir-biridan tarixi, geografik joylashuvi, mahalliy urf-odatlarini, diniy e'tiqodlari va boshqaning farqlari tufayli tarjimada odamlar bir-birini tushunishiga to'sqinlik qiladi. Shunday qilib, tarjima nafaqat tarjimonning lingvistik malakasini balki, tegishli madaniyatlardan xabardor bo'lishini ham inobatga olishi kerak. Shu bilan birga, tarjima faqat so'zlarni, jummalarni yoki maqolalarni manba tildan maqsadli tilga tarjima qilish emas, balki madaniyatlar o'rtasidagi harakatlanishni ham anglatadi⁹⁷. Shu kungacha, milliy o'ziga xos so'zlarni boshqa til vositasida berish sohalarida katta tajribalar to'planib kelmoqda. Badiiy asarlar tarjimasida milliylikni qiyoslaganda asarni boshqa tilga tarjima qilgandagina muammolar paydo bo'ladi.

Bugungi kungacha o'zbek adabiyoti tarixida o'chmas iz qoldirgan hamda uning rivojlanishiga o'zining hissasini qo'shgan buyuk yozuvchilarimiz va shoirlarimiz yetarlicha. Ularning hayot tarzi va boy adabiy merosining yaratilishi va ulardan qolgan badiiy asarlar bugungi kun kitobxonlarini tarix haqida o'ylashga majbur qiladi. Mana shunday asarlardan biri Abdulla Qodiriyning "O'tkan kunlar" romani ham tarixiy roman hisoblanib, o'zbek adabiyoti tarixida yozilgan birinchi romandir. Umuman olganda, Abdulla Qodiriy asarlarining ulug'vor va go'zal bo'lishida badiiy tasvir vositalarining ham o'rni beqiyosdir. Adibning romanlarida tasvirlangan bosh qahramonlar hayotiyiligi, tabiiyligi, qalbning ko'z ilg'amas darajadagi tovlanib turishi, favqulodda kuchli jozibasi, sir asrorlarga boy poetikligi bilan kishini o'ziga maftun etadi.

Ayni paytgacha, asarning rus, nemis, turk, tojik, qozoq, turkman, ingliz va boshqa ko'plab tillarga tarjima variantlari mavjud. Asarni o'zbek tilidan ingliz tiliga bir necha bor tarjima qilishgan, ulardan birinchisi Samarqand Davlat chet tillari instituti rektori Ilhomjon To'xtasinov hisoblanadi. Tarjima qilingan asar Julie Vikenden tomondan tahrirlangan. Romanning ingliz tilidagi talqini Karimov fondi yordami bilan "Nouveau Monde Editions" nomli mashhur fransuz nashriyoti tomonidan nashr etilgan. Asarni rus tilidagi namunasi ingliz tiliga tarjima qilgan adabiyotshunos olim Kerol Ermakovadir⁹⁸. Asarning keyingi tarjimasi AQSh harbiy dengiz akademiyasi mintaqaviy tadqiqotlar markazining sobiq direktori Mark Reese tomonidan ingliz tiliga o'girilgan. Ilmiy tadqiqot davomida aynan mana shu olim tarjimasidagi nusxadan foydalanilgan. Tarjima kitobni o'zbek tahrirchilari Umida Hoshimova va Umida Hikmatillayeva, ingliz tahrirchilari Anne McPeak bilan Sal Robinson va Aleksandria Laureys tomonidan tahrir qilingan. Tanishtiruv(introduction) qismida tarjimon asar haqidagi fikrlarini va Abdulla Qodiriyning "O'tkan kunlar" asarining badiiy va ijtimoiy ahamiyati haqida keltirib o'tgan. Tarjimon asarni tarjima qilish jarayonida Markaziy Osiyoning tarixini va o'ziga xos madaniy so'zlar ma'nosini tushuntirish uchun bir qancha olimlarning asarlaridan foydalangan. Olim uchun kitobni tarjima qilish jarayoni 15 yildan ortiq vaqtni o'z ichiga oldi va 2019-yilda birinchi marta chop qilindi. Quyida asarning o'zbek tilidagi nusxasi va tarjima qilingan shaklidagi ayrim farqlar, shu bilan birgalikda, tarjimonning bu boradagi yutuq va kamchiliklari aniq misollar asosida tasvirlab berilgan.

Asliyatda: *Saroyning to'rida boshqalarg'a qarag'anda ko'rkamrak bir hujra, anovi hujralarga kiyiz to'shalgani holda bu hujrada qip-qizil gilam, uttalarda bo'z ko'rpalar ko'rilgan bo'lsa, munda ipak va **adras** ko'rpalar*⁹⁹...

Inglizcha: *At the far end of the main yard, a room stood apart from the others though its singular beauty: while the other rooms bore only felt rugs, this room displayed deep-red carpets; while others laid their heads on cotton quilts, here the owner spelt among silk and **adras*** bedding;*

Adras* *is a silk-and-cotton-blend fabric. It retains the luster of silk but has a tougher weave to allow for more practical uses – i.e. korpachalar – heavy quilted mattresses with a cotton-filling.*

⁹⁶ Eugene.A.Nida. (2001). Language and Culture-Contexts in Translating. Shanghai: Shanghai Foreign Language Education Press

⁹⁷ Абдуллаева.П.У Таржима жараёнида юзага келадиган маданий муаммолар. Макола, Экономика и сотсиум, (2024). 11-сон(126)

⁹⁸ Choriyeva.M.J "O'tkan kunlar" romanining tarjima variantlarida milliylik ifodasi va o'girmalardagi tafovut va o'xshashliklar. Maqola, "Uchinchi renessans:ilm-fan va ta'lim taraqqiyoti istiqbollari", 2021.

⁹⁹ Abdulla Qodiriy O'tkan kunlar. "O'zbekiston milliy ensiklopediyasi" Davlat ilmiy nashriyoti, Toshkent-2018, 4-bet

*Central Asians use them to sit and sleep on. Every spring the ladies of the household will change out the cotton. These futons are part of the bride's dowry gifted by the groom so they can begin a household*¹⁰⁰.

Ushbu jumlaning tarjima qilish jarayonida “kiyiz” soʻzini “rug” soʻzi bilan berilganini koʻrishimiz mumkin. Vaholanki, kiyiz yaʼni kigiz bu hayvon junidan (odatda qoʻy junidan) tayyorlanadigan zich, issiqni ushlab turuvchi mato. U toʻqilmasdan, junni namlab va bosim ostida zichlab tayyorlanadi. Kigiz koʻproq issiqlikni ushlab turish, uyda yoki chodirda (masalan, qoʻrgʻonda) polga toʻshash uchun ishlatiladi¹⁰¹. “Rug” soʻzi oʻzbek tiliga gilamcha yoki sholcha deb tarjima qilinadi¹⁰². Kigiz va sholcha oʻrtasida maʼno farqlari mavjud. Sholcha yaʼni shol koʻproq yozgi ob-havoda, yerning changidan yoki sovuqligidan saqlanish uchun ishlatiladi¹⁰³. Ayrim hollarda ovqat pishiriladigan joylarda, bogʻ-rogʻlarda ham ishlatiladi. Kigiz ogʻirroq, qalin va silliq yuzali boʻladi. Sholcha esa yengilroq, baʼzida panjara-panjara yoki sargʻish rangda, oʻsimlik tolalaridan qilingan boʻladi. Kigiz maʼnosini ifodalash uchun ingliz tilidagi “felt” soʻzidan foydalanilsa, maʼno aniqroq boʻlar edi.

Keyingi satrlarda esa adras soʻzi oʻzgartirilmasdan berilgan, sababi ingliz tilida bunday mato turi nomini ifodalovchi soʻz yoʻq. Tarjimon esa ushbu soʻzni belgi ostiga olib, maʼnosini sharhlab bergan. Bu esa boshqa madaniyat oʻquvchilari uchun tushunarli boʻlishini taʼminlagan

Asliyatda: Otabek ularga **yuqoridan** joy koʻrsatar ekan, yoqimlilik bir vaziyatda:

*Tinchsizlamadingizlar, bilʼaks quvontirdingizlar, - dedi*¹⁰⁴...

Inglizcha: Otabek appeared unperturbed by the disruption and motioned for them to sit in **the place of honor***. “You do not bother me in the least bit; on the contrary, you lighten my heart with unexpected visit”¹⁰⁵.

The place of honor* - Central Asian gatherings are held in cercles. The Tor, or place of honor, goes to the most senior guest and faces the door from the other side of the room, an area less exposed to drafts and most secure. Since groups sit in a circle, the host should sit opposite the guest or guests, and a junior member of the household, near the doorway, to receive food, placing it in the center of the circle – most notably plov.

Keltirilgan gapda Otabek mehmonlarni yuqoriga yaʼni toʻrga chorlagan, biroq ingliz madaniyatida mehmonni yuqoriga oʻtirgizish yoki toʻr tushunchalari boʻlmagani sababli tarjimon taglama orqali ajoyib tarzda toʻr tushunchasini taʼriflab oʻtgan. Taʼrifda yozilganidek, toʻrga oʻtkazish bu Markaziy Osiyo mehmon kutish madaniyatiga xos jarayon boʻlib, unda toʻrda eng yoshi ulugʻ mehmon oʻtiradi. Bu joy eng kam bezovta qilinadigan va hotirjam joy hisoblanadi. Dumaloq davrada oʻtirilganda mehmonlar bir-birlariga qarama-qarshi holda joy olishadi. Eshik yoniga oʻtirgan uyning eng yosh aʼzosi taomlarni kutib oladi va dasturxon oʻrtasiga uzatadi, bu taom odatda palov boʻladi.

Asliyatda: Otabek mehmonlarni **tanchaga** oʻtquzib, fotihadan soʻng Hasanidan soʻradi: - Tuzukmisiz, ota?¹⁰⁶

Inglizcha: After reciting the **Fatiha*** for the guests sitting around the coal **sandal stove***, Otabek asked Hasan Ali, “Are you well, Father?”*¹⁰⁷

Bu gapda “tancha” yaʼni “sandal” soʻzi keltirilgan. Tarjimada esa bu soʻz “sandal stove” deb tarjima qilingan va taglama asosida maʼnosi quyidagicha tushuntirib berilgan:

Sandal: a Persian-style, low-lying table covered with korpachalar and holding a small atove underneath to keep those seated warm.

¹⁰⁰ Mark Reese Bygone days(Abdulla Qodiriy Oʻtkan kunlar). “Yosh kuch”, Toshkent-2024, 26-bet

¹⁰¹ Akobirov.S.F, Aliqulov T.A va boshqalar. Oʻzbek tilining izohli lugʻati.-Moskva “Rus tili” nashriyoti:1981,1-tom, 385-bet.

¹⁰² Butayev.Sh inglizcha-oʻzbekcha oʻzbekcha-inglizcha lugʻat. Toshkent-2013, 504-bet

¹⁰³ Akobirov.S.F, Aliqulov T.A va boshqalar. Oʻzbek tilining izohli lugʻati.-Moskva “Rus tili” nashriyoti:1981,2-tom, 424-bet.

¹⁰⁴ Abdulla Qodiriy Oʻtkan kunlar. “Oʻzbekiston milliy ensiklopediyasi” Davlat ilmiy nashriyoti, Toshkent-2018, 5-bet

¹⁰⁵ Mark Reese Bygone days(Abdulla Qodiriy Oʻtkan kunlar). “Yosh kuch”, Toshkent-2024, 27-bet

¹⁰⁶ Abdulla Qodiriy Oʻtkan kunlar. “Oʻzbekiston milliy ensiklopediyasi” Davlat ilmiy nashriyoti, Toshkent-2018, 5-bet

¹⁰⁷ Mark Reese Bygone days(Abdulla Qodiriy Oʻtkan kunlar). “Yosh kuch”, Toshkent-2024, 28-bet

Bundan tashqari, bu gapda Otabek Hasanali ga hurmat yuzasidan ota deb murojaat qiladi. Buni ingliz tiliga to'g'ridan-to'g'ri "father" deb tarjima qilingan va ota so'zi faqat o'z otasiga nisbatan emas boshqa holatda ham ishlatilishi mumkinligini izoh orqali berib o'tilgan.

Asliyatda: Hasanali dasturxon yozib qumg'on kirgizdi. Odatiy takalluflar bilan dasturxon va choyga qaraldi. Homid nonni shinniga bulg'ar ekan so'radi:

-Yoshingiz nechada, bek?¹⁰⁸

Inglizchada: Hasan Ali entered with a teapot, spread out the tablecloth, and offered a basin so that the men could wash their hands*. With the time-honored exchange of compliments, inquiries as to one another's health, whether they found themselves too bored, how life treated them, and various other sundry pleasantries, the host began the custom of attending to their guests' needs. Hamid absentmindedly mopped up shinni* with a piece of bread and asked, "How old are you, Bek"¹⁰⁹.

Yuqoridagi gaplarning tarjimasini o'qir ekanmiz, tarjimonning o'zbek madaniyatini qanchalik yaxshi o'rganganligini ko'rishimiz mumkin. "Odatiy takalluflar bilan dasturxon va choyga qaraldi." gapini yozuvchi o'zbeklar odatda choy ustida nimalar haqida suhbatlashishi, choydan oldin mezbon tomonidan kirgizilgan suvda qo'l yuvilishi kabi an'anaviy jarayonlarni kiritib o'tgan. Shu bilan birga, "shinni" va "bek" so'zlarining lug'aviy ma'nosiga ham taglama asosida to'liq izoh berilgan. Bu esa tarjimonning chuqur izlanganligi va mahoratidan deyishimiz mumkin.

Xulosa

Shunday qilib, tarjima bu murakkab tuzilishga ega jarayon hisoblanib, yaxshi tarjima matni chiqishi uchun tarjimondan ancha bilim, malaka va izlanish talab etiladi. Chunki yuqoridagi misollardan ko'rganingizdek, bir millatning milliy qadriyatlariga doir so'zlarni boshqa madaniyat kitobxonlariga tushunarli qilib tarjima qilish har qanday chet tilini biladigan insonning ham qo'lidan kelavermaydi. Yaxshi tarjima chiqishi uchun esa tarjimon nafaqat tilni yaxshi bilishi balki, lingvistika, madaniyatshunoslik, lingvokulturologiya kabi sohalaridan xabardor bo'lishi ham kerak. Agar milliy qadriyatga oid so'z tarjimasida mazmun jihatdan xato yuzaaga kelsa, bu o'zga tilli kitobxon uchun bu millat haqida noto'g'ri taassurot paydo bo'lishiga sabab bo'ladi. Zero, tarjima va madaniyat tushunchalari bir-biri bilan chambarchas bog'liqdir.

Foydalanilgan adabiyotlar:

1. Abdulla Qodiriy O'tkan kunlar. "O'zbekiston milliy ensiklopediyasi" Davlat ilmiy nashriyoti, Toshkent-2018.
2. Абдуллаева.П.У Таржима жараёнида юзага келадиган маданий муаммолар. Мақола, Экономика и сотсиум, (2024). 11-сон(126)
3. Akobirov.S.F, Aliqulov T.A va boshqalar. O'zbek tilining izohli lug'ati.-Moskva "Rus tili" nashriyoti:1981,1-2-tom.
4. Butayev.Sh inglizcha-o'zbekcha o'zbekcha-inglizcha lug'at. Toshkent-2013.
5. Eugene.A.Nida. (2001). Language and Culture-Contexts in Translating. Shanghai: Shanghai Foreign Language Education Press
6. Mark Reese Bygone days (Abdulla Qodiriy O'tkan kunlar). "Yosh kuch", Toshkent-2024.
7. Choriyeva.M.J "O'tkan kunlar" romanining tarjima variantlarida milliylik ifodasi va o'girmalardagi tafovut va o'xshashliklar. Maqola, "Uchinchi renessans:ilm-fan va ta'lim taraqqiyoti istiqbollari", 2021.

¹⁰⁸ Abdulla Qodiriy O'tkan kunlar. "O'zbekiston milliy ensiklopediyasi" Davlat ilmiy nashriyoti, Toshkent-2018, 7-bet

¹⁰⁹ Mark Reese Bygone days(Abdulla Qodiriy O'tkan kunlar). "Yosh kuch", Toshkent-2024, 30-bet

LINGVISTIK TADQIQOTLARDA YANGI YONDASHUV: OSKAR UAYLDNING “DORIAN GREYNING PORTRETI” HIKOYASI ASOSIDA YARATILGAN PARALLEL KORPUS VA UNING TAVSIFI

Bozorqulova O‘g‘iloy Erkin qizi

Toshkent davlat o‘zbek tili va adabiyoti universiteti magistranti

E-mail: ogiloybozorqulova62@gmail.com

Annotatsiya: Ushbu maqolada korpus lingvistikasi sohasining muhim yo‘nalishlaridan biri bo‘lgan parallel korpus va uning tilshunoslikdagi ahamiyati haqida batafsil so‘z yuritiladi. Lingvistika va adabiyot orasida mavjud ilmiy izlanish hamda muammolarga innovatsion yondashuv orqali yechim taklif etiladi. Maqola Oskar Uayldning “Dorian Greyning portreti” asari asosida Sketch Engine dasturi yordamida yaratilgan ingliz-o‘zbek va o‘zbek-ingliz parallel korpusining xususiyatlariga bag‘ishlanadi. Asarning har ikki korpusda ifodalangan til imkoniyatlari, muallif hamda tarjimon qalamiga xos til va uslubiy xususiyatlar qiyosiy tahlil qilinadi.

Abstract: This article talks about the parallel corpus in detail. The parallel corpus is an important area of corpus linguistics, and it is important in linguistics. It offers a new way to understand the connections between linguistics and literature. The article talks about the features of the English-Uzbek and Uzbek-English parallel corpus created using the Sketch Engine program based on Oscar Wilde's work "The Picture of Dorian Gray." The text's linguistic features, including its style and the way it is written, are compared to the text's content.

Keywords: Corpus linguistics, corpus, parallel corpus, thesaurus, e-liner system, n-grams, concordance, syntactic correspondence.

Kirish. Keyingi yillarda lingvistika va texnologiya sohasi o‘rtasida turli tadqiqotlar amalga oshirilishi ko‘payib bormoqda. Bu esa tildagi sintaktik, semantik, lingvokulturologik xususiyatlarni yanada aniq va batafsil ochilishiga, keng ommaga tanishtirilishiga turtki bo‘lmoqda. Shuningdek, korpus lingvistikasi sohasida ham bir qancha amaliy ahamiyatga ega ishlar olib borilmoqda. Bu yangiliklardan biri parallel korpusning yaratilishi va uning lingvistik izlanishlarda keng foydalanishning yo‘lga qo‘yilishi bo‘ldi.

Parallel korpus – biror tilning matnlarini boshqa tilga tarjima qilingan matnlar bilan o‘z ichiga olgan maxsus ma’lumotlar bazasi sanaladi. Bunday korpus turi tilshunoslarga nafaqat tarjima jarayonini tahlil qilish, balki tilning leksik, sintaktik, morfologik va semantik xususiyatlarini chuqur o‘rganish imkonini yaratadi. Parallel korpuslar, asosan, ma’lum bir kontekstdagi turli til tuzilmalari, iboralar va so‘zlarning tarjimasini ko‘rish uchun ishlatiladi[1]. Parallel korpuslarni tuzishda bir qancha qiyinchiliklar mavjud bo‘lib, bulardan matn tozalash jarayoni alohida o‘rinda turadi. Sababi ikki tilli parallel korpuslarni tuzish jarayonida asl matn (source language) va tarjima matnlarini (target language) moslashtirish (alignment) jarayoni juda muhim bosqich hisoblanadi. Ushbu jarayon keyingi tahlil jarayonlarining sifatli va aniq chiqishiga yordam beradi. Yana shuni ta’kidlash joizki, mashina tarjimasida so‘zma-so‘z tarjima qilish o‘zini oqlamagan usul bo‘lib, katta qatorni tashkil etuvchi murakkab gaplar tarjima matnda kichik hajmdagi jumalalarga o‘zgarishi mumkin. Shu sababdan parallel matnlarni tanlashda ushbu xususiyatlarni nazarda tutish lozim. Badiiy asar bo‘yicha parallel matnlar korpusi, avvalo, ikki yoki bir necha tillar grammatikasidagi o‘xshash va farqli jihatlarni tahlil qilish imkonini beradi.. Parallel matnlar korpusi har ikki tildagi badiiy uslub xususiyatlarini qiyoslashda ahamiyatlidir[2]. Parallel korpuslar badiiy asarlarning asl nusxasi va tarjima qilingan variantlarini birlashtirish orqali tarjimadagi o‘zgarishlarni tahlil qilish, shuningdek, til va madaniyatlar o‘rtasidagi farqlarni aniqlash imkonini beradi. Ushbu korpuslar, ayniqsa, leksik va sintaktik strukturalar qiyosini

amalgam oshirishda muhim ahamiyatga ega bo'lib, tarjima jarayonida yuzaga keladigan o'ziga xosliklarni chuqur o'rganishga zamin yaratadi.

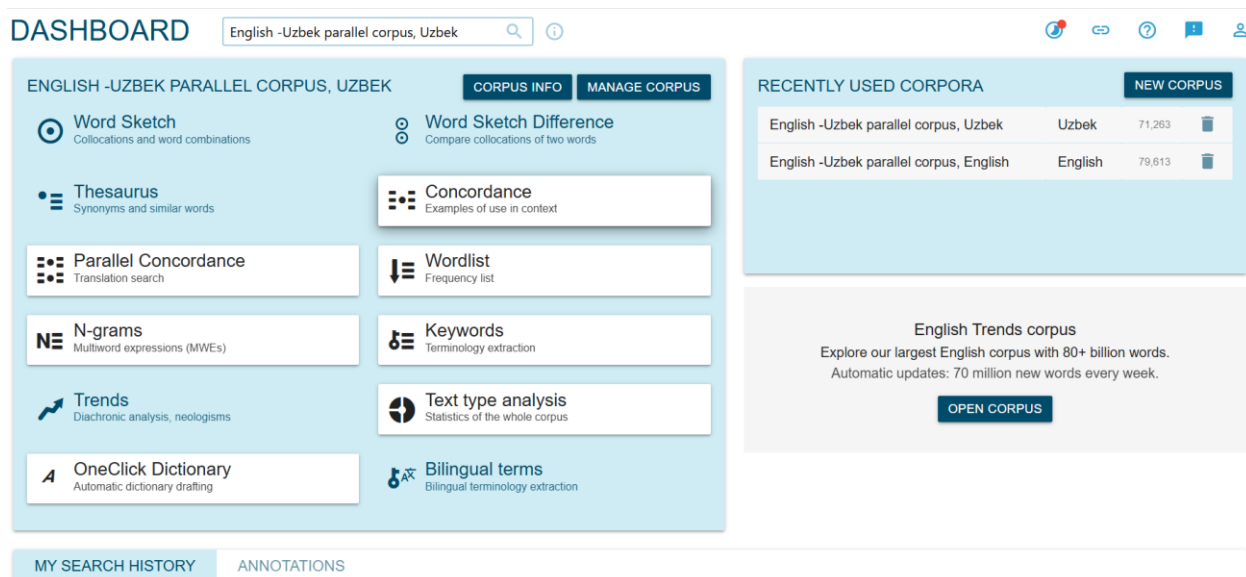
Muhokama va natijalar

Quyida Oskar Uayldning “Dorian Greyning portreti” hikoyasi asosida tuzilgan ingliz-o'zbek va o'zbek-ingliz parallel korpusidan sitatistik, qiyosiy, lingvistik tahlillar jadvallari asosida beriladi. Va ushbu matnlar o'zaro moslashtirilib Sketch Engine dasturida avtomatik alingment jarayonidan o'tdi va bunda moslashtirishning eng samarali usuli bo'lgan gaplarni muvofiqlashtirish (sentence alignment) usulidan foydalanildi. Hikoyaning Ozod Sharafiddinov tomonidan o'zbek tiliga tarjima qilingan varianti tanlab olindi. Korpus haqidagi statistik ma'lumotlarni quyidagi jadvallarda ko'rish mumkin.

1-jadval. Ingliz-o'zbek hamda o'zbek-ingliz parallel korpusi ma'lumotlari

No		Ingliz-o'zbek PK	O'zbek-ingliz PK
1	Words	112,626	108,920
2	Tokens	158,194	140,759
3	Sentences	8,651	7175
4	Paragraphs	728	782

Ushbu jadvalda bir asarning ikki tildagi nusxalaridagi farqlarni yaqqol ko'rish mumkin. Jumladan, parallel korpusda asar ingliz tilidan o'zbek tiliga tarjima qilinganda o'zbek tilidagi nusxasida mavjud so'zlarning 4000 tadan ko'proqqa qisqarganini ko'rish mumkin. Yoki tarjima nusxada gaplarning qariyb 1500 taga qisqargani ko'zga tashlanadi. Parallel korpus tuzishdan avval asarning ikki tildagi ayni nusxalarining elektron formatlari topildi. Keyin dasturga muvaffaqiyatli yuklanib, ikki tilli parallel korpus ishga tushdi(1-rasm).



1-rasm. Parallel korpus interfeysi.

Asar tahliliga kirishishdan avval kalit so'zlarni topib olish kerak bo'ladi. Quyidagi jadvalda ikki korpusda topilgan dastlabki 15 ta kalit so'zlarni ko'rish mumkin(2-jadval).

O'zbek-rus PK	Rus-o'zbek PK

2-jadval. Parallel korpusdagi kalit soʻzlarning qiyosiy tahlili.

Kalit soʻzlar	Takrori	Baholash	Kalit soʻz	Takrori	Baho
Dorian	345	1148.97	Dorian	650	4555.61
lord	105	935.314	hallword	554	3922.5
Bezil	139	684.623	Sibly	430	3055.87
Sibila	107	626.397	Vane	276	1961.45
yoʻq	115	626.039	narborough	283	1851.61
Genri	110	578.605	Henry	258	1816.63
Grey	119	535.356	Gray	200	1421.87
Xolluord	155	528.829	handsome	176	1251.14
Gersoginya	113	523.543	duchess	172	1222.95
boʻlsa	86	523.337	hideous	120	834.737
goʻzal	75	475.101	dreadfull	137	804.019
surat	70	395.237	marvellous	91	631.597
koʻzlar	282	389.307	mysterious	85	598.323
gʻaroyib	58	367.638	listlessly	78	555.139
gʻalati	67	365.094	horribly	87	509.835

Ushbu jadvalni taqqoslaganda ingliz-oʻzbek va oʻzbek-ingliz parallel korpusidagi kalit soʻzlar tarjimada biroz oʻzgarganini koʻrish mumkin. Ammo tanlagan asarimiz murakkab strukturalar va madaniy tafovutga ega soʻzlarga boy boʻlmagani uchun ikkala asardagi kalit soʻzlar bir-biridan katta farq qilmaydi. Ushbu qiyosiy tahlildan yana bir xulosaga kelish mumkin boʻladi. Asarning tarjima nusxasida kishilarning lavozim, mansab hamda laqablari oʻzbek tilidan toʻgʻridan-toʻgʻri olib kirilgan.

Parallel korpusda badiiy asar tahliliga oid muhim jihat prallel konkordansning mavjudligidir. Har ikki asarning gaplar muvofiqligi-yu, ikki tilga xos boʻlgan leksik tuslanishlar ushbu tahlil jarayonida yuzaga chiqadi. Bunda parallel korpus kalit soʻzlarida yuzaga chiqqan soʻzlarni tanlasak, tahlil yana-da chuqurroq amalga oshiriladi. Masalan, “sanʼat” soʻzini ushbu tahlil jarayonida qidirib koʻrdik. Bunda biz konkordans lugʻatda uchta tahlilni amalga oshirdik. Birinchi tahlil, matnlar korpusda toʻgʻri muvofiqlashtirilganligini guvohi boʻldik. (2-rasm).

PARALLEL CONCORDANCE

type to search

simple art • 9
97.7 per million tokens • 0.0098%

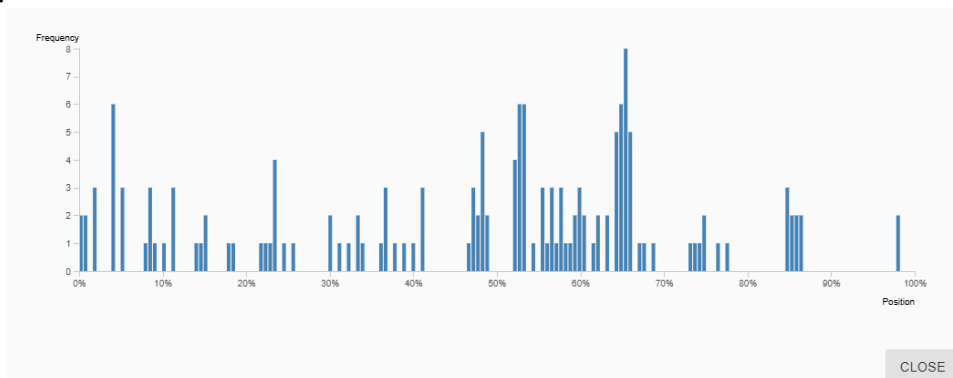
English -Uzbek parallel corpus, Uzbek

doc#0	<s> To reveal art and conceal the artist is art's aim. </s>	<s> Sanʼat intiladigan maqsad — odamlarga oʻzini oshkor qilib, ularning nazaridan sanʼatkorini yashirib qolishdir. </s>
doc#0	<s> To reveal art and conceal the artist is art's aim. </s>	<s> Sanʼat intiladigan maqsad — odamlarga oʻzini oshkor qilib, ularning nazaridan sanʼatkorini yashirib qolishdir. </s>
doc#0	<s> The moral life of man forms part of the subject- matter of the artist, but the morality of art consists in the perfect use of an imperfect medium. </s>	<s> Sanʼat axloqi esa — nomukammal vositalarni mukammal qoʻllashdan iboratdir. </s>
doc#0	<s> Thought and language are to the artist instruments of an art. </s>	<s> Fikr va Soʻz sanʼatkor uchun sanʼat vositasidir. </s>
doc#0	<s> It is the spectator, and not life, that art really mirrors. </s>	<s> Mohiyatan, Sanʼat shunday bir koʻzgudirki, unda sira ham hayot aks etmaydi, balki unga qaragan odamning aksigina jilolanadi. </s>
doc#0	<s> Diversity of opinion about a work of art shows that the work is new, complex, and vital. </s>	<s> Agar sanʼat asari bahslarga sabab boʻlsa, demak, unda allanechuk yangi, murakkab va muhim narsa bor. </s>
doc#0	<s> All art is quite useless. </s>	<s> Har qanday sanʼat mutlaqoʻ befoyddir. </s>
doc#0	<s> The first is the appearance of a new medium for art, and the second is the appearance of a new personality for art also. </s>	<s> Bizning davrimizda odamlar sanʼat asari avtobiografiyaga oʻxshagan bir narsa boʻlmogʻi kerak deb oʻylashadi. </s>
doc#0	<s> The first is the appearance of a new medium for art, and the second is the appearance of a new personality for art also. </s>	<s> Bizning davrimizda odamlar sanʼat asari avtobiografiyaga oʻxshagan bir narsa boʻlmogʻi kerak deb oʻylashadi. </s>

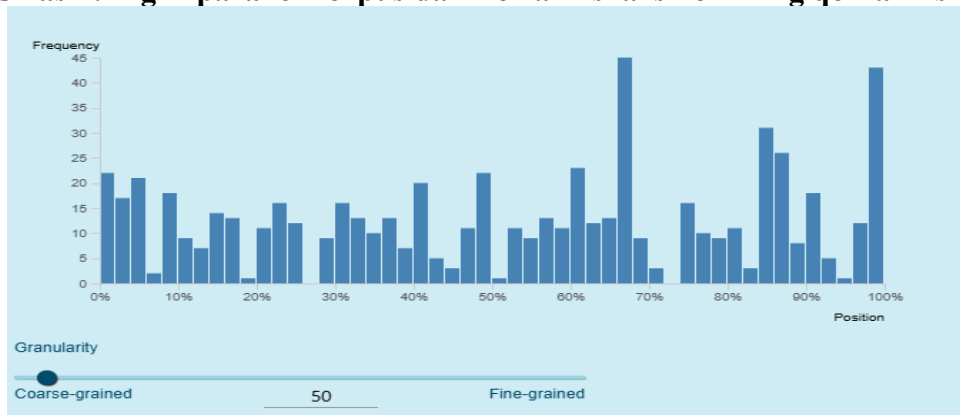
2-rasm. Matnlar muvofiqlashtirish(alignment) parallel korpus.

Ikkinchi qiyosiy tahlil “Dorian” shaxs nomining ikki korpusda ham asarda qoʻllanilganlik darajasini diagrammada ifodalanishi boʻldi(3-rasm). Ikki korpus diagrammasida ham koʻrish mumkinki, Asarning 70 % qismiga yaqinlashganda har ikki asarda ham “Dorian” ismidan eng

yuqori darajada foydalanilgan. Va, aksincha, o‘zbek-ingliz parallel korpusida ushbu ismga asarning 85% foizidan boshlab oxirigacha ko‘p takrorda murojaat qilingan bo‘lsa(4-rasm), ingliz-o‘zbek parallel korpusida tarjima asarning 85% dan to oxirigacha bo‘lgan qismda umuman tilga olinmagan.



3-rasm. Ingliz parallel korpusida “Dorian” shaxs nomining qo‘llanilishi.



4-rasm. O‘zbek tilli parallel korpusda “Dorian” shaxs nomining qo‘llanilishi.

Keyingi qiyosiy tahlil parallel korpusda ingliz tili va o‘zbek tilidagi so‘zlarni qidirish imkoniyatlarini ko‘rsatib berdi. Masalan, har ikki tilli korpusda ham qidiruvni yakka, fraza, so‘z va hokazo shakllar ustida amalga oshirish mumkinligi ko‘rinadi. Biroq keyingi bosqichlarda rus tili parallel korpusining imkoniyatlari kengayadi. Ya’ni qidiruvdagi birlikning so‘z turkumini, o‘rnini aniqlash mumkin bo‘ladi. O‘zbek tilidagi asar korpusida esa so‘z turkumlarini belgilash imkoni mavjud emasligi ayon bo‘ldi(5-rasm).

CONCORDANCE English -Uzbek parallel corpus, English

BASIC ADVANCED LEARN

Query type ①

- simple
- lemma
- phrase
- word**
- character
- CQL

Part of speech

- any
- adjective
- adverb
- conjunction
- determiner
- noun**
- numeral
- particle

Word

women

✓ A = a ?

Subcorpus ① none (the whole corpus)

Macro ? none

Filter context ②

Text types ?

GO

5-rasm. Parallel konkordans qidiruv imkoniyatlari.

Keyingi tahlilni amalga oshirish ham parallel korpus imkoniyatlarini ko‘rsatib beruvchi muhim jihatdir. Bu N-grammlar asosida matnda qo‘llanilgan turli leksik birliklarning bog‘lanish

shakllarini topish imkoni mavjud. Badiiy asarning lingvistik mohiyati ham ushbu tahlilda yuzaga chiqadi. Masalan, “Dorian Greyning portreti” asarining asl tilida 2-grammlardan foydalanib ko’rdik. Bunda eng ko’p qo’llanilgan so’z juftliklari, asosan, “*Lord Genri*”, “*Dorian Grey*”, “*san’at*”, “*surat*”, “*ayollar*” bo’ldi.

O‘zbek tili PK	Ingiliz tili PK

3-jadval. 2-gramm asosida qiyosiy tahlil.

2-gramm natijasi	1037 ta	1367 ta
Takrori	12 690 marta	14 613 marta

Xulosa. Oskar Uayldning “Dorian Greyning portreti” hikoyasi asosida yaratilgan ingliz-o‘zbek parallel korpusida statistik, qiyosiy, lingvistik tahlillarni amalga oshirib, quyidagi xulosalarni qayd etish mumkin bo’ldi:

1. Oskar Uayldning “Dorian Greyning portreti” hikoyasi o‘zbek tilidagi varianti korpusga joylanganda so‘zlar soni asliyat tildagisidan 4000 taga, gaplar miqdori esa qariyb 1000 taga kamaygan. Bunda tarjima asarda ba’zi gaplar va jumlar qisqartirilgan.

2. Asarning asliyat tilida 2-grammlar tahlili asosida shu aniq bo‘ldiki, kishilarning nomlaridan keyingi o‘rinda keng qo‘llanilgan leksik birliklar qo‘shma olmoshlar, ko‘makchili birikmalardir.

3. Shaxs nomlari tarjima asarda o‘zbek tilidagisidan 3 barobar ko‘p qo‘llanilgan(masalan, “Dorian” so‘zi 450 marta). Bu esa tarjimonga xos xususiyatlardan bir ekanligi namoyon bo‘lgan.

4. Ingliz-o‘zbek parallel korpusida parallel konkordans, tezaus funksiyalari shuni ko‘rsatdiki, ingliz tilidagi matnlar tahlilida korpus imkoniyatlari kengayadi. Buni o‘zbek tilida korpus lingvistikasi hamda parallel korpus yangi soha ekani, hali taraqqiy etmagani hamda “Sketch Engine” dasturida ko‘p tilli korpus yaratish imkoniyatlarining samaradorligini oshirish zarurligi bilan izohlaymiz.

Foydalanilgan adabiyotlar:

1. Abjalova M. (2022). Korpus lingvistikasi. Toshkent, 42-bet.
2. M.Akramdjanova. (2023). Parallel matnlar korpusini yaratishning amaliy va nazariy ahamiyati / Ilm-fan va innovatsiya ilmiy-amaliy konferensiyasi(in-academy.uz/index.php/si). Tashkent.
3. Abjalova M. Parallel korpuslarni yaratish asoslari (2023). Kompyuter lingvistikasining zamonaviy texnologiyalari-CTCL.2023. –Tashkent, 35-42-bet.
4. Mengliyev B, Hamroyeva Sh. (2020). Korpus lingvistikasi: Korpus tuzish va undan foydalanish. Tashkent.
5. Sobirova N, Akramdjanova M, Sobirov Sh. (2022). Parallel korpus tuzish xususiyatlari, tuzilishi , interfeysi / Academic research in Educational Sciences, 34-bet.
6. https://app.sketchengine.eu/#ngrams?corpname=user%2Fmushtariyalimjonova7%2Fenglish_uzbek_parallel_corpus_uzbek&tab=basic&ngrams_n=2&ngrams_max_n

O‘ZBEK VA INGLIZ NUTQIY ETIKETIDA GENDERGA ASOSLANGAN KOMMUNIKATIV FARQLAR

Kurbanova Muhabbat Matyakubovna

O‘zbekiston Milliy universiteti professori, f.f.d

E-mail: qurbonova2007@mail.ru

Bultakova Durdona Shokirovna

O‘zbekiston Milliy universiteti tayanch doktoranti

E-mail: bultakovadurdona0@gmail.com

Annotatsiya: Ushbu maqolada o‘zbek va ingliz tillarida nutqiy etiket tizimi doirasida genderga xos kommunikativ farqlar tahlil qilinadi. Muloqot jarayonida erkak va ayol suhbatdoshlarning til birliklarini tanlashi, xushmuomalalik vositalaridan foydalanishi, nonverbal harakatlar va ijtimoiy maqomga nisbatan munosabati o‘zaro solishtiriladi.

Kalit so‘zlar: nutqiy etiket, gender, sotsiopragmatika, muloqot, madaniyatlararo farq, nonverbal vosita, til va jins.

Abstract: This article analyzes gender-based communicative differences within the speech etiquette systems of Uzbek and English languages. It compares the choice of linguistic units, politeness strategies, use of nonverbal signals, and the reflection of social hierarchy in conversations between male and female speakers.

Keywords: speech etiquette, gender, sociopragmatics, communication, intercultural difference, nonverbal cues, language and gender.

Аннотация: В статье анализируются гендерные коммуникативные различия в системе речевого этикета в узбекском и английском языках. В процессе общения сравнивается выбор языковых единиц, использование вежливости, невербальных действий и отношение к социальному статусу собеседниками мужского и женского пола.

Ключевые слова: речевой этикет, гендер, социопрагматика, коммуникация, межкультурные различия, невербальные средства коммуникации, язык и пол

Bugungi kun tilshunosligi til va uning foydalanuvchisini markazga qo‘ygan holda ularni turli aspektlarda o‘rganib, ilm-fan uchun yangi xulosalar bermoqda. Xususan, o‘zbek va ingliz tillari grammatik qurilishi, tarixiy taraqqiyoti, rivojlanish yo‘nalishlari, tarqalishi, ichki va tashqi omillarning ta’siri, tildan foydalanuvchilarning soni, tilning leksik, fonetik, sintaktik qurilishi, shuningdek, nutqiy faoliyatning pragmatik, lingvokulturologik, sotsiologik jihatlar bilan farqlanadi. Nutqiy etiketni o‘rganish aspektlaridan biri, uni har ikki tilda genderologik jihatdan tadqiq etishdir. XX asrning oxiri XXI asrning boshlarida tilshunoslikda markazida inson turuvchi bir qator yangi yo‘nalishlar: sotsiolingvistika, pragmalingvistika, lingvokulturologiya, ekolingvistika, biolingvistika, kognitiv lingvistika, neyrolingvistika, psixolingvistika kabi sohalar qatorida ayol va erkak nutqining xususiyatlarini batafsil yorituvchi genderolingvistika ham taraqqiy etdi. Tarixan jamiyat ijtimoiy faoliyatida bola tarbiyalovchi va uy-ro‘zg‘or ishlari bilan shug‘ullanuvchi vazifasida bo‘lgan ayollarning nutqiy faoliyati, jamiyat faoliyatining ijtimoiy boshqaruvchisi bo‘lgan erkaklarnikidan sezarli darajada farqlanishi ularning muloqot etiketida ma’lum farqlarni yuzaga keltiradi. Tildan foydalanuvchilarning ijtimoiy holati, kasb-kori, jamiyatdagi mavqeyi, yoshi, ma’naviy salohiyati, shaxsiy mansubligi, yashash hududi kabi omillar bilan bir qatorda so‘zlovchilarning jinsi ham ular nutqining ayrim farqli jihatlarining namoyon bo‘lishida muhim ahamiyatga ega hisoblanadi.

Gender tushunchasi zamonaviy tilshunoslik paradigmasiga boshqa gumanitar fanlar bilan taqqoslaganda kechroq kirib kelgan bo‘lib, o‘tgan asrning 60-yillarida atama sifatida shakllana

boshlandi. Gender tushunchasi ilm-fanda birinchi marta amerikalik psixanalitik J.Stoller tomonidan qo'llanilgan. Olim 1986-yilda "Gender: a Useful Category of Historical Analysis" ("Gender: tarixiy tahlilning foydali kategoriyasi") nomli maqolasini ommaga e'lon qildi¹¹⁰. Olim gender atamasini biologik va sotsiologik tushuncha sifatida talqin qilgan. J.Stollerning olib borgan ilmiy izlanishlari zamonaviy gumanitar fanlar orasida gender tilshunosligi deb nomlangan alohida yo'nalishning rivojlanishiga olib keldi. Shuningdek, A.Brandt gender – madaniyat ta'siri ostida bo'ladigan, ma'naviy jihatdan ayol va erkaklarning nutqiy, xulq-atvoriy tomonlarini baholay oladigan xususiyatlar majmuyidir, deb qaraydi¹¹¹. Dunyo madaniyatlarida shaxsning xulq-atvori o'lchovlari odatda jinsga nisbatan farqli talablarni o'z ichiga oladi. Bu jamiyat talab etgan ijtimoiy qonuniyatlarga ayollar va erkaklarning rioya qilishlarini bildiradi. Tilshunos olim Y.Goroshko gender xususiyatlar nutq, xulq-atvor, ayol va erkaklarning jamiyat modellarini qabul qilishda o'z aksini ko'rsatadi, deb ta'rif etadi. J.Lakoffning ta'kidlashicha, ayollar nutqiy etiketi yuqori darajadagi hamkorlik, suhbatdoshiga moslashib ketish va uning fikr-mulohazalarini diqqat bilan eshitish, muloqotda ustunlikni qo'lga olishga harakat qilmaslik kabi xususiyatlari bilan ajralib turadi.

Gender tadqiqotlari jinslar o'rtasidagi farq va munosabatlarni o'rgansa, feminizm masalalariga bag'ishlangan ishlarda esa xotin qizlar muammosini tadqiq etishga asosiy urg'u beriladi¹¹². Aslida, gender germeneytikasi "xotin-qizlarga taalluqlilik" degan mazmundan ancha kengroq tushunchadir. Chunki "xotin-qizlarga taalluqlilik" jinslararo xulq, muloqot shakllariga taalluqli bo'lgan, jinsiy rollarni bajarishni o'rganishga qaratilgan tadqiqotlar va modellardan tashkil topgan. Genderda (tadqiqot obyekti sifatida) har ikkala jinsning bir-biri bilan munosabati, aloqadorligi va turli darajadagi ijtimoiy tizimlar doirasida (oilal, ishlab chiqarish birlashmasi va boshqalarda) o'zaro determinatsiyasi kabi masalalar qamrab olingan bo'ladi. Gender farqlar jinslarning biologik farqlanishi, ya'ni tabaqalashishi natijasi emas, balki ularning ijtimoiylashuvi va tarbiyasining jamiyatda qaror topgan tasavvur hamda xulq modeliga amal qilishi natijasidir.

Tahlil natijalari ayollar va erkaklar o'rtasidagi nutqiy farqlar nafaqat til darajasida, balki madaniyatga xos kommunikativ qadriyatlar orqali shakllanishini ko'rsatdi.

- O'zbek nutqiy etiketida, ayollar ko'proq muloyim, hurmatga asoslangan iboralardan foydalanadilar ("Jon aka", "Yaxshi qoling, opa", "Iltimos, uzr bo'lsin"), erkaklar esa rasmiy va buyruq ohangidagi uslublarga moyil ("Keling", "Bo'ldi", "Qani"). Ayollar ko'pincha ijtimoiy mavqeni hisobga olib muomala qiladilar.

- Ingliz nutqida ayollar "hedging" (yumshatuvchi so'zlar), "politeness markers" (xushmuomalalik belgilar) ko'p qo'llaydilar: "Would you mind...?", "Sorry to bother you...". Erkaklar esa qisqaroq, to'g'ridan-to'g'ri iboralarni tanlaydilar ("Let's do it", "I'll take care of it").

- Diskursiv tahlil shuni ko'rsatdiki, ayollar muloqotda hamkorlikni kuchliroq ifoda etadi, ya'ni savollar berish, tasdiqlovchi iboralar, tasalli beruvchi ifodalarni afzal ko'radilar. Erkaklar esa ko'proq axborotga yo'naltirilgan, bahsli yoki ustunlikni bildiruvchi uslublardan foydalanadilar.

O'zbek nutqiy etiketida ayollar turmush o'rtog'iga uning ismi bilan murojaat qilishi tabu hisoblanib, uning o'rniga "dadas", "otasi", "u kishi", "anovi" kabi so'zlarni qo'llaydi. Shuningdek, erkaklar ham rafiqasining ismi bilan murojaat qilishi holatlari kam uchraydi. Erkaklar katta farzandining ismini aytish orqali rafiqasiga murojaat qilishi ayrim hududlarimizda odat tusiga kirgan.

Ayollar nutqiy etiketining o'ziga xos jihatlari, ayniqsa, leksikasi va intonatsiyasida yaqqol namoyon bo'ladi. Shu bilan bir qatorda o'zbek ayollari nutqiga xos "girgitton", "aylanay", "o'rgilay", "qoqindiq" kabi so'zlar ijobiy ma'no nozikligini ifodalovchi leksik vositalar sanaladi. Shuningdek, ushbu so'zlar ko'proq keksa ayollar muloqotida faol qo'llaniladi: 1. *Qo'shnilar ahvol*

¹¹⁰ Brown C. Understanding Other Cultures. – New Jersey: Prentice-Hall, 1963. – P. 354.

¹¹¹ Потопов. V. Многоуровневая стратегия в лингвистической гендерологии // Вопросы языкознания, 2002. – С. 110-130.

¹¹² Zubin D. Gender and noun classification // International encyclopedia of linguistics. Vol. 2. – Oxford: Oxford University Press. 1992. – P. 41-43.

so‘rasa, ularniyam, o‘ziniyam yupatadi: – “Ha, endi keksalikda, o‘rgilay, lekin onamning oyoq og‘rig‘i faqat keksalikdan emas. 2. – Kel-a, bo‘yningda girgitton bo‘lib ketay, seni qaysi shamol uchirdi, akam tirilib keldimi, haligina qovog‘im uchib turgan edi, – dedi ammam”. Ijobiy ma‘no nozikligini ifodalovchi leksik vositalar ingliz ayollari nutqini bezovchi birliklar sanaladi. Ular yaqin insonlariga “sweetie”, “honey”, “darling”, “dear”, “gravy” deya murojaat qilishadi. Aynan shu so‘zlar o‘zbek ayollari tomonidan ham qo‘llanadi. Inglizlar kichkina bolalarga “sweetie” deb murojaat qilishsa, o‘zbek ayollari “asolim” deb aytishadi. Inglizlar turmush o‘rtog‘i yoki rafiqasiga “darling” deb murojaat qilishsa, o‘zbeklar “azizim”, “azizam”, “begim” deya murojaat qilishadi. Ayollar muloqot etiketi salomlashish va so‘rashish jarayonining uzoq davom etishi, ayni fikrlarni takrorlash, hissiyotlarni ochiq va ravon ifodalash, boshqalarning gaplarini suhbatdoshiga yetkazish, tanishlarining gaplaridan suhbat davomida foydalanish, suhbat chog‘ida ma‘noli so‘zlarning qo‘llanishi, fikr-mulohazalarni mubolag‘a qilish orqali ifodalash, muloqot paytida pauzalarning deyarli bo‘lmasligi, gaplarning uzun bo‘lishi, suhbatdoshning fikrini tez-tez bo‘lishi kabi xususiyatlari bilan ajralib turadi. O‘zbek va ingliz xalqlari muloqot xulqidagi o‘xshash va farqli jihatlar murojaat shakllarida yaqqol ko‘zga tashlanadi. Qolaversa, biz bu jihatlarini oilada ayolning o‘zini tutishi, turmush o‘rtog‘iga munosabatida ham ko‘rishimiz mumkin. O‘zbek va ingliz madaniyatlarida ayollar nutqida xushmuomalalik, yumshoqlik va ijtimoiy me‘yorlarga bo‘ysunish kuchliroq ifodalanadi. Ayollar iltimos, minnatdorchilik va uzr so‘rashda ko‘proq ijtimoiy maqbul strategiyalarni tanlaydi. Erkaklar muloqotda ko‘proq rasmiylikka, aniq va qisqa ifodalarga moyildirlar. Bu holat ikkala madaniyatda ham kuzatiladi, biroq O‘zbek tilida erkaklar nutqida ba‘zan buyruq va to‘g‘ridan-to‘g‘ri murojaatlar keng tarqalgan. Ingliz tilidagi gender kommunikatsiyasi o‘zbek tiliga nisbatan ancha nozik, diplomatik va kontekstga bog‘liq bo‘lib, ijtimoiy tenglik g‘oyalariga yaqinlashgan. O‘zbek nutqiy etiketida erkak va ayol rollari ijtimoiy stereotiplarga ko‘proq bog‘langan. Shu sababli ayollarning nutqi ko‘proq ijtimoiy konvensiyalar bilan cheklangan bo‘lishi mumkin.

Ayollar va erkaklar nutqiy etiketini qat‘iy chegaralab bo‘lmaydi. Shu bilan birga, genderlik xususiyatlarining ma‘lum soni ayollar nutqiga va boshqalari erkaklar nutqiga xos deb aytish qisman noo‘rin. Chunki ayollar va erkaklarning lisoniy nutq shakllarini tanlashiga turli omillar, ya‘ni har bir shaxsning etnik kelib chiqishi, irqiy tegishliligi, dini, yoshi, kasbi kabilar ta‘sir qiladi. Badiiy adabiyot tahlillarida ayollar va erkaklar orasidagi muloqot xususiyatlari quyidagilarda aks etgan: a) badiiy adabiyotlarda ayol va erkakning jamiyatdagi o‘rni; b) ayol va erkak nutqining fonetik jihatdan qo‘llanilishi; d) ayol va erkak nutqining grammatik jihatdan qo‘llanilishi; e) ayol va erkak nutqida leksik vositalarning qo‘llanilishi¹¹³. Ingliz badiiy adabiyotidan olingan ba‘zi misollarda ayol dominantasi yaqqol seziladigan, shuningdek, erkak ayolning fikrini rad eta olmaydigan holatlar uchraydi. Ingliz madaniyatida bunday holat odatiy hisoblanib, bu o‘zbek lingvomadaniyatida deyarli uchramaydi. Ingliz badiiy adabiyotidan olingan misollarda ayollar muloqot etiketidagi e‘tiborga molik yana bir xususiyatning guvohi bo‘ldik. Bunda ayollar nutqida suhbatdoshiga nisbatan hurmat, sabrlik, bir-birini o‘zaro tushunishga intilish strategiyalari namoyon bo‘ladi. Har ikki tilda genderga asoslangan kommunikativ farqlar pragmatik kompetensiyani shakllantirishda muhim omil hisoblanadi, ayniqsa chet tilini o‘rganishda ushbu farqlarni anglash va hisobga olish til o‘zlashtirish sifatini oshiradi.

Foydalanilgan adabiyotlar:

1. Atadjanova N. Muloqot jarayonida lisoniy mansublik tavsifi va turlari. Filologiya fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi avtoreferati. – Qo‘qon, 2022. – 26 b.
2. Brown, P. & Levinson, S. (1987). Politeness: Some Universals in Language Usage. Cambridge University Press.
3. Karimova M. “Nutq madaniyati asoslari”. – Toshkent: O‘qituvchi, 2005. – B. 74.
4. Yo‘ldoshev O. “Sotsiopragmatik tilshunoslikka kirish”. – Toshkent: Fan, 2010. – B. 89.

¹¹³ Гендер тадқиқоти асослари курси хрестоматияси. –Тошкент: “Ўзбекистон”, 2003. – Б. 264.

5. Ҳакимов М. Ўзбек тилида матнинг прагматик талқини. Филол. фан. д-ри ...дисс.. – Тошкент, 2001. – 283 б.
6. Калилина Г. Речевой этикет в разных структурированных языках. Дис. док... филол. наук. – Москва. 2021. – 214 с.
7. Мўминов С. Ўзбек мулоқот ҳуққининг ижтимоий-қисмий хусусиятлари. Филол. фан. номзоди. дис.– Тошкент, 2000. – 235 б.
8. Zubin D. Gender and noun classification // International encyclopedia of linguistics. Volume- 2. Oxford: Oxford University Press. 1992. – P. 41-43.
9. Тухтасинов И. Лингвокультурологические и гендерные особенности сложных слов в художественном тексте: на материале английского и узбекского языков: Дисс. канд. филол. наук. – Ташкент, 2011. – 160 с.

SEMANTIK QIDIRUV ALGORITMI

Fayzullayeva Z.I.

Muhammad al-Xoraymiy nomidagi TATU dotsent (PhD)

E-mail: zarnigor18z02@gmail.com

Abdurashidova O.A.

Toshkent davlat Sharqshunoslik universiteti magistranti,

E-mail: abdurashidov0809@icloud.com

Zoidjonova O.M.

Toshkent davlat Sharqshunoslik universiteti magistranti,

E-mail: oydinoyzoidjonova@gmail.com

Annotatsiya: Ushbu tadqiqot semantik qidiruv texnologiyalarini va ularni SEO strategiyalariga qo'llashni o'rganadi. Tadqiqotning asosiy maqsadi internet resurslarini qidiruv tizimlariga moslashtirish uchun yagona matematik model ishlab chiqishdir. Tadqiqotda semantik qidiruvning asosiy komponentlari, jumladan, matnni vektorlashtirish, so'zlarni normallashtirish, stop-so'zlarni chiqarib tashlash va semantik og'irlikni hisoblash kabi jarayonlar tahlil qilinadi. Shuningdek, semantik qidiruvga mos keluvchi SEO tavsiyalari, ayniqsa kontent yaratish, savol-javob shaklida matnlar tuzish va ontologik yondashuvlar keltirilgan. Tadqiqot natijalari semantik qidiruv algoritmlari asosida optimallashtirilgan kontentni avtomatik ishlab chiqish imkonini beradi. Tadqiqotda semantik qidiruvni qo'llash orqali qidiruv tizimlarida yuqori natijalarga erishish uchun zarur bo'lgan metodologiya va texnik tavsiyalar ishlab chiqilgan. Kelajakda bu tadqiqotning natijalari, ayniqsa, yangi tillar va lingvistik bazalar uchun lokalizatsiyalangan modullar yaratishda qo'llanilishi mumkin.

Kalit so'zlar: SEO, semantic qidiruv, vektorlash, normallashtirish, stop-word.

Annotation: This research explores semantic search technologies and their application to SEO strategies. The primary goal of the study is to develop a unified mathematical model for optimizing internet resources for search engines. The research analyzes the key components of semantic search, including text vectorization, word normalization, stop-word removal, and semantic weight calculation. Additionally, the study provides SEO recommendations that align with semantic search, particularly in content creation, question-answering formatted texts, and ontological approaches. The research results enable the automatic generation of content optimized according to semantic search algorithms. The findings of the study offer methodologies and technical recommendations for achieving high rankings in search engine results through semantic search. In the future, these results can be applied, especially in the creation of localized modules for new languages and linguistic databases.

Keywords: SEO, semantic search, vectorization, normalization, stop-word.

Аннотация: Это исследование рассматривает технологии семантического поиска и их применение в SEO-стратегиях. Основная цель работы — разработать единая математическую модель для оптимизации интернет-ресурсов для поисковых систем. Исследование анализирует ключевые компоненты семантического поиска, такие как векторизация текста, нормализация слов, удаление стоп-слов и расчет семантического веса. Кроме того, в работе приведены рекомендации по SEO, соответствующие семантическому поиску, в частности, в создании контента, формировании текстов в формате вопрос-ответ и онтологических подходах. Результаты исследования позволяют автоматически генерировать контент, оптимизированный с учетом алгоритмов семантического поиска.

Выводы исследования предлагают методологии и технические рекомендации для достижения высоких позиций в результатах поисковых систем с использованием семантического поиска. В будущем эти результаты могут быть применены, особенно при разработке локализованных модулей для новых языков и лингвистических баз данных.

Ключевые слова: SEO, семантический поиск, векторизация, нормализация, стоп-слово.

Internet resurslarining soni yildan-yilga keskin ortib borayotgani fonida foydalanuvchilarga kerakli va mazmunli axborotni topish tobora murakkab holga kelmoqda. An'anaviy qidiruv tizimlari, asosan, kalit so'zlar asosida ishlaydi va matn mazmunining chuqur semantik tahlilini amalga oshirmaydi. Bu esa sinonimlar, ko'p ma'nolilik (polisemiya) va kontekst muammolari tufayli ko'plab nomuvofiq natijalar paydo bo'lishiga olib keladi. Semantik qidiruv texnologiyalari ushbu muammolarning oldini olishda samarali vosita bo'lib xizmat qilmoqda. Ontologiyalardan foydalanish, foydalanuvchi niyatini aniqlash, matnlar o'rtasidagi semantik o'xshashlikni hisobga olish bular semantik qidiruvning asosiy ustunliklaridandir. Mazkur tadqiqotda internet resurslarini semantik qidiruv texnologiyalariga mos ravishda optimallashtirish uchun yagona matematik apparat — algoritmlar algebrasi asosida modellashtirish yondashuvi taklif etiladi. Ushbu yondashuv orqali kontentni baholash, o'xshashlik darajasini aniqlash, va keyinchalik ontologiyalarni qurish imkonini beruvchi algoritmik model ishlab chiqiladi.

So'nggi yillarda axborotni izlash texnologiyalarida sezilarli o'zgarishlar kuzatildi. Klassik qidiruv mexanizmlari asosan Boolean qidiruv (AND, OR, NOT operatorlari), yulduzcha bilan ifodalgan umumiy belgilar (wildcard), masofa bo'yicha qidiruv (proximity search) va noaniq qidiruv (fuzzy search) kabi usullarga asoslangan. Biroq, bu yondashuvlar foydalanuvchi ehtiyojini to'liq qondirishda zaiflik qiladi, chunki ular so'zlarning kontekstual ma'nosini hisobga olmaydi [1,4,6]. Masalan, kontekstual qidiruv – foydalanuvchi yozgan so'zning kontekstini aniqlashga harakat qiladi. Bu yondashuv "Crystal Semantics" tizimi asosida ishlovchi Textonomy texnologiyasida namoyon bo'lgan. Ushbu yondashuv foydalanuvchi so'zining boshqa so'zlar bilan bog'liqligini aniqlab, ma'noli natijalarni qaytaradi [10]. Bugungi kunda semantik qidiruv — ontologiyalar, semantik tarmoqlar, va entitetlar (ya'ni shaxslar, joylar, hodisalar kabi obyektlar) asosida quriladi. Masalan, foydalanuvchi "Lviv" va keyin "universitet" deb izlaganida, tizim foydalanuvchining "Lvivdagi universitetlar" haqida ma'lumot izlayotganini tushunadi.

Bunday tizimlar Latent Semantic Indexing (LSI), Latent Dirichlet Allocation (LDA) va TF-IDF (Term Frequency-Inverse Document Frequency) kabi statistik va semantik modellar orqali amalga oshiriladi. Ushbu usullar matnlar o'rtasidagi o'xshashlik darajasini aniqlash, asosiy so'zlarni ajratib olish, va ularga semantik vazn berishda qo'llaniladi [14,20]. Biroq, bu usullar hali to'liq universal emas va ko'p hollarda tillarga bog'liq, lokalizatsiya masalalari bor. Shu sababli, algoritmlar algebrasi asosida yondashuvni taklif etish, ayniqsa turli tillarda (jumladan, o'zbek va ukrain tillarida) barqaror modellashtirishga zamin yaratadi [3,13].

Tadqiqotning asosiy vazifalari

Ushbu tadqiqotning asosiy maqsadi – semantik qidiruv texnologiyalaridan foydalangan holda internet resurslarini qidiruv tizimlariga moslashtirish uchun yagona matematik model yaratishdir. Bunday yondashuv axborot resurslarining mazmunini baholash, semantik moslikni aniqlash va qidiruv tizimlari bilan integratsiyalashgan tarzda takomillashgan SEO strategiyalarini ishlab chiqish imkonini beradi. Tadqiqot quyidagi asosiy vazifalarni o'z ichiga oladi:

- Mavjud semantik qidiruv texnologiyalarini tahlil qilish. Foydalanuvchi ehtiyojlarini aniqlashda, turli tillarda ishlov berish muammolarini hal etishda qo'llanilayotgan algoritmlarni ko'rib chiqish.
- Algoritmlar algebrasi asosida matematik modelni sintez qilish. Ushbu yondashuv yordamida matnlarni vektor ko'rinishida tasvirlash, semantik o'xshashlikni aniqlash va ontologiyalarni qurish uchun zamonaviy matematik asos yaratish.

- Modellashtirilgan model asosida SEO tavsiyalarini ishlab chiqish. Kontentni semantik jihatdan boyitish, foydalanuvchi niyatiga mos kontent ishlab chiqish va qidiruv tizimlari uchun optimallashtirish strategiyalarini ishlab chiqish.
- Semantik baholash uchun maxsus algoritmlarni yaratish. Ushbu algoritmlar orqali matnlar o'rtasida TF-IDF, LSI, LDA kabi mezonlar asosida o'xshashlik darajasini hisoblash mumkin bo'ladi.
- Ontologik model tuzish orqali semantik bog'liqlikni aniqlash. Mazmuniy birliklar (klasslar, atributlar, aloqalar) asosida RDF va OWL formatlarida ontologiyalar yaratish va ularni semantik qidiruvga moslashtirish.

Asosiy natijalar va modellashtirish

Semantik qidiruv foydalanuvchi ehtiyojlarini nafaqat so'zlar darajasida, balki ularning kontekstual va semantik ma'nolari orqali aniqlashga qaratilgan. Ushbu texnologiyada matnni raqamli ko'rinishga keltirish, ortiqcha (informatsion yuklamaga ega bo'lmagan) so'zlarni chiqarib tashlash, asosiy so'zlarni ajratish va semantik ontologiyalar tuzish kabi bir necha bosqichli yondashuv mavjud.

Matnni formula orqali ifodalash uchun v vector ko'rinishiga o'tkaziladi

$$F(v) = [t_1, t_2, \dots, t_n]$$

Bu yerda t_i - matndagi leksik birliklar

So'zlarni normallashtirish uchun ortiqcha qo'shimchalar va morfemalar olib tashlanadi. Bu jarayon esa quyidagicha ifodalanadi

$$F(s) = \text{stem} \min g(F(v))$$

So'zlar orasidagi semantik ahamiyati quyidagi formula orqali hisoblanadi

$$TF - IDF(t, d) = TF(t, d) \times \log \left(\frac{N}{DF(t)} \right)$$

$TF(t, d)$ - hujjatda nechi marta uchrashi

$DF(t)$ - so'z uchraydigan hujjatlar soni

N - umumiy hujjatlar soni

Kalit so'zlarni ajratib olish uchun esa quyidagi formuladan foydalaniladi

$$F(kL) = \text{KeywordList} F(TF - IDF)$$

Yuqoridagi formulalardan kelib chiqqan holda umumiy algoritmnı quyidagicha bo'ladi

$$F_{SEO} = F(o) \circ F(kL) \circ F(mS) \circ F(d) \circ F(s) \circ F(v)$$

Xulosa

Ushbu tadqiqot davomida semantik qidiruv texnologiyalaridan foydalangan holda internet resurslarini optimallashtirishning matematik modeli ishlab chiqildi. An'anaviy qidiruv tizimlarining imkoniyatlari semantik murakkabliklar (masalan, sinonimiya, polisemiya, kontekstli bog'liqlik) oldida cheklanganligi isbotlandi. Algoritmlar algebrasi asosida yagona matematik model sintez qilindi. Bu model matnni qayta ishlash, stemming, stop-so'zlarni chiqarib tashlash, semantik og'irlikni hisoblash va ontologiya yaratishni o'z ichiga oladi. Semantik tahlil algoritmi taklif qilindi. Ushbu algoritm yordamida kontentlar o'rtasida o'xshashlik darajasi aniqlanadi va foydalanuvchi niyatiga mos natijalarni shakllantirish mumkin. Ushbu tadqiqot natijalari semantik qidiruvga mos keluvchi zamonaviy SEO platformalarini yaratish, ularni avtomatlashtirish va tarmoqli tizimlarga integratsiya qilish imkonini beradi

Foydalanilgan adabiyotlar:

1. Grappone, J. Search Engine Optimization (SEO): An Hour a Day. Wiley Publishing, 2013.

2. Su, J. va boshq. Model of Touristic Information Resources Integration According to User Needs. CSIT, 2018.
3. Ovsyak, V. Algoritmlar: tuzilish, optimallashtirish va ehtimollik asoslari. Lviv, Svit, 2001.
4. Basyuk, T. Veb-saytni targ'ib qilish va anchor-so'zsiz optimallashtirish. CSIT, 2016.
5. Basyuk, T. Ichki bog'lanishlar va havola og'irligi. CSIT, 2017.
6. Amerland, D. Google Semantic Search. Que Publishing, 2013.
7. Vysotska, V. va boshq. Lingvometrik koeffitsiyentlar yordamida mualliflik aniqlash. CSIT, 2019.
8. Najman, L., Talbot, H. Mathematical Morphology. Wiley-ISTE, 2010.
9. Frank, Y. Shih Image Processing and Mathematical Morphology. CRC Press, 2009.
10. Jones, K. A statistical interpretation of term specificity. Journal of Documentation, 2004.
11. Boshqa CSIT, FQAS va Springer nashrlari. To'liq ro'yxat asosiy maqola asosida tuzilgan.

O'ZBEK TILI ASOSIDA QORAQALPOQ TILINING NLP TIZIMINI YARATISH

Otemisov Aziz Zarliqbaevich

QDU, Amaliy filologiya kafedrasini mudiri,
filologiya fanlari bo'yicha falsafa doktori (PhD), dotsent.
E-mail: utemisov.aziz@mail.ru

Xudaybergenova Gozzal Kenesbay qizi

Qoraqalpoq davlat universiteti
Kompyuter lingvistikasi ta'lim yo'nalishi 2-kurs talabasi
E-mail: Xudaybergenovagozzal670@gmail.com

Annotatsiya: Ushbu maqolada o'zbek va qoraqalpoq tillarida NLP texnologiyalari qiyosiy tahlil qilinadi. Avvalo, o'zbek tilida yaratilgan NLP loyihalari, xususan, UzBERT va BBPOS modellarining imkoniyatlari, lingvistik korpuslar (UZCorpus, Universal Dependencies) va ularning tilda semantik, sintaktik va morfologik darajadagi tahlilga qo'shgan hisssasi ko'rib chiqiladi. Ammo qoraqalpoq tilida hali NLP bo'yicha yetarli ilmiy-texnik resurslar yo'qligi, ayniqsa lemmalangan, teglangan korpuslar va tahlil vositalarining yo'qligi ta'kidlanadi. Shuningdek, parallel korpuslar yaratish, mavjud modellarni fine-tuning qilish, lemmatizator va tokenizatorlarni moslashtirish kabi amaliy yechimlar taklif etiladi..

Kalit so'zlar: NLP texnologiyasi, SI, mashina tarjimai, transfer learning, morfologik tahlil, sintaktik tahlil, parallel korpus, lemmatizator, tokenizator

Аннотация: В статье представлен сравнительный анализ технологий обработки естественного языка (NLP) для узбекского и каракалпакского языков. Прежде всего рассматриваются существующие NLP-проекты для узбекского языка, в частности, потенциал моделей UzBERT и BBPOS, а также вклад лингвистических корпусов (UZCorpus, Universal Dependencies) в семантический, синтаксический и морфологический анализ. Отмечается недостаточность научно-технических ресурсов по NLP для каракалпакского языка, особенно в части лемматизированных и размеченных корпусов, а также аналитических инструментов. Кроме того предлагаются практические решения, включая создание параллельных корпусов, дообучение (fine-tuning) существующих моделей, а также адаптацию лемматизаторов и токенизаторов.

Ключевые слова: обработка естественного языка (NLP), искусственный интеллект (ИИ), машинный перевод, перенос обучения (transfer learning), морфологический анализ, синтаксический анализ, параллельный корпус, лемматизатор, токенизатор.

Abstract: This article presents a comparative analysis of natural language processing (NLP) technologies for the Uzbek and Karakalpak languages. First, it examines existing NLP projects for the Uzbek language, particularly the potential of models such as UzBERT and BBPOS, as well as the contribution of linguistic corpora (UZCorpus, Universal Dependencies) to semantic, syntactic, and morphological analysis. The study highlights the lack of scientific and technical resources for NLP in the Karakalpak language, especially in terms of lemmatized and annotated corpora and analytical tools. Furthermore, practical solutions are proposed, including the creation of parallel corpora, fine-tuning of existing models, and the adaptation of lemmatizers and tokenizers.

Keywords: natural language processing (NLP), artificial intelligence (AI), machine translation, transfer learning, morphological analysis, syntactic analysis, parallel corpus, lemmatizer, tokenizer.

Tabiiy tilni qayta ishlash (NLP) - bu mashinalarga inson tilini o'qish, shifrlash, tushunish va qayta ishlash imkonini beruvchi eng mashhur SI texnologiyasi hamda inson tili ma'lumotlarining yozilishi, og'zaki aytilishi va tartibga solinishi bilan boshqara oladigan mashinalarni yaratishdir. Matnni taxmin qilish va his-tuyg'ularni tahlil qilishdan tortib, nutqni aniqlashgacha, NLP mashinalarga inson aqli va qobiliyatlarini ta'sirchan tarzda takrorlash imkonini beradi. NLP ilovalarini yaratish uchun til modellari kalit hisoblanadi. Biroq, noldan murakkab NLP til modellarini yaratish ancha qiyin, shuning uchun SI va ML ishlab chiquvchilari va tadqiqotchilari oldindan tayyorlangan til modellari bilan hamkorlikda ishlaydi. Til modellari nutqni tanish, belgilarni optik aniqlash, qo'lyozmani aniqlash, mashina tarjimasini va imloni to'g'rilash kabi NLPning turli masalalarini yechish uchun qo'llaniladi.

Tabiiy tilni qayta ishlash (NLP) texnologiyalari, so'nggi yillarda, sun'iy intellektning rivojlanishi bilan birga, turli tillarning raqamli tahlilini soddalashtirib kelmoqda. Ayniqsa, o'zbek va qoraqalpoq tillari uchun NLPning rivojlanishi, bu tillarning morfologik o'xshashliklari va resurslarining mavjudligi nuqtai nazaridan juda muhim.

O'zbek tilida NLP bo'yicha bir qator ishlar amalga oshirilgan. 2021 yilda Mansurov tomonidan ishlab chiqilgan UzBERT modeli, o'zbek tilida tilning semantik va sintaktik jihatlarini qayta ishlashda yuqori natijalarga erishgan. UzBERT, BERT (Bidirectional Encoder Representations from Transformers) asosida ishlab chiqilgan model bo'lib, o'zbek tilidagi matnlarni yaxshiroq tushunish va ularga mos ravishda javob berishni ta'minlaydi. Shu bilan birga, BBPOS [Bobojonov, 4-12] modeli o'zbek tilidagi so'z turkumlarini aniqlashda NLPga muhim hissa qo'shgan.

O'zbek tilidagi NLP ishlanmalari, korpuslar va lingvistik resurslar. O'zbek tili uchun yaratilgan korpuslar tabiiy tilni qayta ishlash (NLP) jarayonining tayanch manbalaridan biri hisoblanadi. UZCorpus va Universal Dependencies Uzbek Treebank [Devlin, Chang, Lee, Toutanova, 1-6] kabi loyihalar til strukturasi, sintaktik aloqa va semantik tahlil uchun zarur bo'lgan asosiy manbalarni taqdim etadi. Bunday korpuslar o'zbek tilining morfologik boyliklarini hisobga olgan holda mashinali o'rganish modellarini samarali tayyorlash imkonini beradi.

O'zbek tilining NLP uchun mos kelish afzalliklari. O'zbek tilining fonetik yozuvga asoslanganligi, morfologik qat'iylik va so'z yasash tizimining izchilligi NLP tizimlarini ishlab chiqishda muhim afzalliklar beradi. Til strukturasi aniq va qat'iy bo'lgani sababli, mashina tomonidan tahlil qilish uchun mos algoritmlarni ishlab chiqish nisbatan osonlashadi. Shuningdek, so'nggi yillarda yaratilgan lingvistik korpuslar va model loyihalari UzBERT [Mansurov, 1-7], BBPOS [Bobojonov, 99-104] o'zbek tilida NLP tadqiqotlarini yanada jadallashtirmoqda.

Qoraqalpoq tilidagi NLP holati: Qoraqalpoq tili uchun tabiiy tilni qayta ishlash (NLP) sohasida hozircha yetarli darajada ilmiy-amaliy qo'llanmalar yo'q. Bu holat, birinchi navbatda, til korpuslarining cheklanganligi yoki mavjud emasligi bilan bog'liq. Zamonaviy NLP texnologiyalari uchun zarur bo'lgan lemmalashtirilgan, morfologik belgilangan va sintaktik tahlil qilingan matnlar to'plami qoraqalpoq tilida deyarli ishlab chiqilmagan. Shuningdek, elektron lug'atlar, so'z turkumlarini belgilovchi belgilangan ma'lumotlar bazalari, so'zlar o'rtasidagi morfologik va semantik aloqalarni tahlil qiladigan modellar yetishmaydi. Bu esa qoraqalpoq tilida avtomatik tarjima, emotsional tahlil, nutqni aniqlash kabi tizimlarning rivojlanishini sekinlashtirmoqda.

O'zbek va qoraqalpoq tillari turkiy tillar oilasiga kirib, ular o'rtasida leksik, morfologik va sintaktik jihatdan yaqinlik mavjud. Bu o'xshashliklar, ayniqsa, so'z yasash, gap tuzilishi va grammatik qoidalarda yaqqol ko'zga tashlanadi. Bu yaqinlik, o'zbek tilida yaratilgan NLP modellarini qoraqalpoq tiliga moslashtirishda asosiy omil bo'lib xizmat qilishi mumkin.

O'zbek tilida mavjud korpuslar, masalan, UZCorpus, o'zbek tilining morfologik va sintaktik xususiyatlarini aks ettiradi. Bu korpuslar, tarjima tahlil orqali qoraqalpoq tiliga moslashtirilishi mumkin. Tarjima tahlil, parallel korpuslar yaratish orqali amalga oshiriladi, bu esa qoraqalpoq tilida NLP modellarini yaratishda muhim manba bo'lib xizmat qiladi.

"Cross-lingual transfer learning" (ko'p tili o'rganish) [Devlin, Chang, Lee, Toutanova, 4171-4179] usuli yuqori resursli tillarda yaratilgan NLP modellarini kam resursli tillarga moslashtirish imkonini beradi. Transformer arxitekturasida yaratilgan modellar, masalan, mBERT (multilingual BERT), ko'p tili ma'lumotlar asosida o'rgatilgan bo'lib, ularning yordamida qoraqalpoq tilida NLP vazifalarini bajarish mumkin. Bu usul, ayniqsa, qoraqalpoq tilida mavjud ma'lumotlar cheklangan bo'lsa-da, ijobiy natijalar berishi mumkin.

Qoraqalpoq tilida matnlar ko'p, lekin internet orqali qidirib topish imkoniyati cheklanganligi sababli, o'zbek tilidagi matnlarni tarjima qilish orqali yangi korpuslar yaratish mumkin. Bu jarayonda, parallel korpuslar tuzish orqali qoraqalpoq tilida NLP modellarini o'rgatish uchun zarur bo'lgan ma'lumotlar bazasi yaratiladi. Bu usul, qoraqalpoq tilida NLP tadqiqotlarini rivojlantirishda muhim qadam bo'lib xizmat qiladi.

Qoraqalpoq tilida tabiiy tilni qayta ishlash (NLP) uchun zarur bo'lgan lingvistik va texnik resurslar - masalan, elektron korpuslar, teglangan matnlar, morfologik analizatorlar va sintaktik daraxtbanklar hozirgi kunda deyarli yo'q. Bu holat, qoraqalpoq tilida NLP loyihalarini yo'lga qo'yishda katta to'siq bo'lib kelmoqda.

Bu muammoni bartaraf etishda o'zbek tilida mavjud bo'lgan NLP resurslarini - ayniqsa, UZCorpus, o'zbek tili morfologik analizatori, lemmatizatorlar va tokenizatorlarni moslashtirish orqali foydalanish mumkin. Chunki o'zbek va qoraqalpoq tillari morfologik va sintaktik jihatdan juda yaqin bo'lib, bu modellarni minimal o'zgarishlar bilan qoraqalpoq tiliga o'tkazish imkonini beradi. Shuningdek, parallel korpuslar yaratish orqali ikki til o'rtasidagi leksik va grammatik mosliklarni aniqlab, mavjud o'zbek modellarini transfer learning yoki fine-tuning orqali qoraqalpoq tiliga moslashtirish mumkin. Bunday kirish, kam resursli tillarda NLPni rivojlantirishda keng qo'llanilmoqda va qoraqalpoq tilida ham muvaffaqiyatli natijalar berishi mumkin.

Shuningdek, parallel korpuslar yaratish orqali ikki til o'rtasidagi leksik va grammatik mosliklarni aniqlab, mavjud o'zbek modellarini transfer learning yoki fine-tuning orqali Qoraqalpoq tiliga moslashtirish mumkin [Tideman, 2214-2218].

Xulosa qilib aytganda, o'zbek va qoraqalpoq tillari morfologik va sintaktik jihatdan yaqin bo'lganligi sababli, o'zbek tilida mavjud bo'lgan NLP resurslari qoraqalpoq tilida ham muvaffaqiyatli qo'llanilishi mumkin. Hozirgi vaqtda qoraqalpoq tili uchun NLP tizimlari yetarli darajada rivojlanmagan bo'lsa-da, o'zbek tilidagi mavjud modellardan foydalanib, parallel korpuslar yaratish va transfer learning usullarini joriy etish orqali bu sohada tez va samarali rivojlanishga erishish mumkin. Bu munosabat qoraqalpoq tilining raqamli maydonda saqlanishi va rivojlanishiga xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Mansurov Sh. (2021). UzBERT: Pretraining a BERT model for Uzbek. arXiv:2108.09814.
2. Bobojonova M. va boshqalar. (2023). BBPOS: O'zbek tilida so'z turkumlarini aniqlovchi neyron tarmoq modeli.
3. Mamasaidov I., Shopulatov Z. (2022). Open Language Data for Low-Resource Turkic Languages.
4. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
5. Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. In Proceedings of LREC 2012.
6. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
7. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).

MANIFESTATION OF THE LINGUISTIC PERSONALITY OF A POLITICIAN IN THE UZBEK LANGUAGE

Usmonova Lola Komil qizi

National University of Uzbekistan

E-mail: lola.usmonova94@mail.ru

Annotation This article analyzes the manifestation of the linguistic personality of a politician, its linguistic and communicative properties. Through the means of political speech and rhetoric, the methods of forming the personality, influence on society and image of the politician are illuminated. The study explores the semantic, pragmatic, linguocultural and stylistic aspects of political discourse and analyzes the mechanisms of their influence on the audience.

Keywords: political speech, semantics, pragmatics, stylistics.

Аннотация: В данной статье анализируется проявление языковой личности политика, ее лингвистические и коммуникативные свойства. Средствами политической речи и риторики освещаются методы формирования личности, влияния на общество и имиджа политика. В исследовании рассматриваются семантические, прагматические, лингвокультурные и стилистические аспекты политического дискурса и анализируются механизмы их влияния на аудиторию.

Ключевые слова: политическая речь, семантика, прагматика, стилистика.

In the current process of globalization, the manifestation of politicians as a linguistic person is gaining importance in the life of society. Political speech is not only a means of communicating information, but also a weapon to direct the masses, form their views and influence social consciousness. The political figure of the politician is manifested through his individual speech characteristics, rhetorical style, communicative strategies and discursive characteristics. In discourse, however, the politician's speech culture stands out. After all, speech culture – involves the use in all areas related to language in accordance with clear, understandable and literary norms. The correctness of speech consists not only in adhering to grammatical rules, but also in choosing expressions that correspond to the colloquial environment. Therefore, the study of the linguistic personality of a politician is considered one of the pressing issues of today's modern linguistics.

The concept of linguistic identity is multifaceted and is studied by areas of linguistics such as pragmalinguistics, Psycholinguistics, sociolinguistics, cognitive linguistics, linguoculturology. At the intersection of these areas in linguistics, a relatively new field linguopersonology was formed. The main subject of study of linguopersonology is a linguistic person, and in linguistics the term began to be widely used since the 90s. The concept was first introduced in 1927 by the German linguist Y. Weisgerber's appearance in his treatise "the formation of the mother tongue and the soul" is recorded in the scientific literature [Usmanova, 2019. 179]. As a linguistic person, in many cases, the ability of a person to create and use speech is recognized. But a linguistic person means not only a person who can create speech and understand it, but also a language owner who knows its grammatical features, can bring the possibilities of expression to the surface, can compose a specific speech [Umurzoqova, 2023. 110-118]. At this point, the linguistic personality is a multifaceted concept, and it is necessary to take into account the variety of factors that also cause it. In the formation of a linguistic personality, in particular, along with the linguistic factor, extralinguistic factors are calculated: the position of a person, profession, craft, position, age, gender, religion, nationality, etc. are of paramount importance. A person is the chief organizer of communication. Speech communication, on the other hand, does not take place outside of man. Speech communication is influenced by various factors related to the individual [Saatova, 2024. 37-38]. The politician's linguistic identity, on the other hand, assumes special attention from linguists as a motivating person who speaks to media, public and international audiences in

connection with such factors. His speech is studied not only as a means of communicating information, but also as a linguistic phenomenon with the power of social influence.

The linguistic personality of a politician is formed through his oratory characteristics, eloquent culture, rhetorical skills and communicative strategies. Linguoculturological factors play an important role in the formation of political speech, since through speech, national mentality, cultural values and traditional speech norms are reflected. The speech characteristics of the leaders operating on the political scene of Uzbekistan are also closely related to the national mentality, cultural values and historical factors. In their speeches, we can see the following aspects: 1) the reflection of national traditions and values – political speeches reflect Uzbek sincerity, folk expressions and traditional values. 2) emphasis on development and reform – the topic of modernization, development and reform is widely covered in the speeches of heads of state and politicians. 3) the promotion of the ideas of unity and solidarity – the speeches will strongly express the ideas of national unity, collective interest and striving for progress. These aspects determine the specificity of the speech of politicians and play an important role in their formation as a linguistic person.

When political speech is linguoculturologically analyzed, the following factors are manifested in the leading position in the political discourse of the Uzbek language: 1. Traditionalism is the reflection of the Customs and values of the Uzbek people (e.g. "*Dear countrymen!*" forms of address such as). 2. National thought-images and images that fit into the historical and cultural context (for example, "*going down the path of our great ancestors...*"). 3. An emotional and moving style is the use of warm, intimate forms of speech to ensure intimacy with the people ("*my dear people...*").

Sh.Rahmatullayev researched the linguistic, psycholinguistic and cultural aspects of Stylistics. He believes that speech culture is not only grammatical correctness, but also includes factors that ensure that speech is rich in content, expressive and purposeful. Uzbek linguists (e.g. Sh. Ruziyev, T. Mirzayev) emphasize the harmony of the lexical, grammatical and phonetic layers of the language in the formation of speech culture. And the manifestation of the culture of speech in the person of each politician is seen in the fact that he tries to convey his worldview to the masses through his individual style of speech, metaphors and impressive expressions. In the process, political speech serves as a means of persuading the audience, attracting their attention and guiding them towards the goal.

A number of studies have been carried out on the formation of a politician as a linguistic person, his oratory characteristics and methods of influencing the masses. According to the results of this study, the speech of politicians has several features from a linguistic point of view. In particular, in the Uzbek language, we can cite the following main features in the manifestation of the linguistic personality of a politician:

1. Lexical-semantic feature-in political speech, affective words, emotionally rich phrases, metaphors and jargons are often used. For example, phrases such as "*in the interests of the people*", "*a confident step towards progress*" are often found in political speech. 2. Syntactic feature - long and complex sentence devices, parallel structures, and rhetorical questions play an important role in increasing the effectiveness of speech. Therefore, in political discourse, such devices are often used, and in order to have a long-term effect, politicians use rhetorical structures that support strong arguments and social values. 3. A pragmatic feature is that in order to convince and influence the audience, politicians use different pragmatic strategies. Examples of such strategies include forms of address ("*dear people*", "*dear compatriots*"), logically based arguments, and speeches reinforced by Statistics.

The main goal of political speech is to convince the audience and form their opinion. The following factors are important in this process: 1) rhetorical strategies – politicians use rhetorical techniques based on emotional and logical arguments to make their speech influential. 2) Public Relations – politicians express their positions and connect with the public through the media, social networks and various performances. 3) the effectiveness of speech-the influence of a politician's

speech on the masses will be directly related to his rhetorical skill, emotionality and relevance [Craft., Cremin, 2014. 16-34].

There have been a number of scientific studies on the personality of the politician Lisoni, one of which is Van Deyk's research on political discourse analysis. Van Deyk's study of political discourse analysis is one of the main sources in understanding political discourse and linguistic strategies. According to the author, political discourse is an important means of influencing society, it serves to influence consciousness through the medium of language, to form certain views among social groups [Lin, 2011. 149]. Van Deyk studies political discourse through cognitive and social aspects and analyzes the basic rhetorical techniques used by policymakers. According to him, political speech acquires a logical structure: politicians in their speeches often use an evidence-based strategy, which makes their words more convincing. Van Deyk's research provides an important scientific basis for how political discourse is used to shape public opinion, as well as to reveal the political power of language.

It can be seen from this that political discourse and discourse are not only a means of conveying information, but also an effective mechanism for the formation of social consciousness and the strengthening of power. Politicians consciously develop speech strategies regularly, seeking to attract the attention of the audience and expand their sphere of influence.

In conclusion, the success of a politician's linguistic personality depends on how he uses speech strategies. Politicians use different linguistic tools to influence their audience, and each of them can lead to certain results. The study shows that the most effective strategies are the creation of emotional impact and a sense of intimacy with the people, which play an important role in attracting the attention of the audience and gaining confidence. In general, the success of a politician's linguistic personality depends on how he communicates with the audience. Effective linguistic strategies and carefully formulated speech serve to strengthen the public image of the politician.

Literature used:

1. Craft A., Cremin T. Creative primary schools: developing and maintaining pedagogy for creativity, *Ethnogr. Educ.*, t.9, 1, 16-34, January 2014
2. Lin S. Fostering creativity through education a conceptual framework of creative pedagogy. *Creat. Educ.* t 2.3., page 149, 2011.
3. Saatova S. O'zbek tilida lisoniy shaxsni shakllantiruvchi ijtimoiy omillar. – filologik tadqiqotlar: til, adabiyot, ta'lim. 2024, 3-son
4. Umurzoqova M. Lisoniy shaxs va uning shakllanishida ijtimoiy omilning o'rni. *International scientific-methodological electronic journal "Foreign Languages in Uzbekistan"*, 2023, No 6 (53), 110-118
5. Usmanova Sh. *Lingvokulturologiya*. – Toshkent, 2019.
6. Курбонова М. Ўзбек болалар нутқининг прагматик хусусиятлари: Филол. фан. докт. ... дисс. – Тошкент, 2018.– 240 б.
7. Ниязова Д. Бадиий матнда лисоний шахс типлари: Филол. фан. бўйича фалсафа докт. дисс. – Қарши, 2020. – Б. 51.

SO‘Z SHAKLLARINING ASOSINI ANIQLASH ALGORITMI

Fayzullayeva Z.I.

Muhammad al-Xorazmiy nomidagi TATU dotsent

E-mail: zarnigor18z02@gmail.com

Karimov N.N.

Namangan davlat universiteti o'qituvchisi,

E-mail: topcoder1600@gmail.com

Abdumutalov B.M.

Muhammad al-Xorazmiy nomidagi TATU talaba

E-mail: student1105@tuit.uz

Rahmatullayev O.O.

Muhammad al-Xorazmiy nomidagi TATU talaba

E-mail: orifjonrahmatullayev3@gmail.com

Annottasiya: So‘z shakllarining asosini aniqlash (ildiz shaklini topish) tabiiy tilni qayta ishlash (NLP) sohasida muhim ahamiyatga ega bo‘lib, bu jarayon morfologik tahlilning asosini tashkil etadi. So‘z shakllarini aniqlash orqali grammatik va semantik xususiyatlar aniqlanadi, bu esa tilni to‘g‘ri tushunish va texnologiyalarni ishlab chiqishda yordam beradi. Ushbu jarayon ma'lumotni tasniflash, tarjima qilish, muloqot tizimlarini yaratish kabi NLP dasturlarida zarurdir. So‘z shakllarini aniqlash uchun turli metodlar mavjud, jumladan, qoidalarga asoslangan yondashuvlar, morfologik modullar, lemmatizatsiya, stemming, mashina o‘rganish va chuqur o‘rganish usullari. Har bir yondashuvning o‘ziga xos afzalliklari va kamchiliklari bor. Masalan, qoidalarga asoslangan yondashuv oddiy va samarali bo‘lsa-da, murakkab grammatik strukturalarni to‘liq aks ettira olmaydi. Morfologik modullar esa yuqori aniqlikni ta'minlaydi, lekin katta ma'lumotlarga ehtiyoj bor. Lemmatizatsiya va stemming metodlari tilda nozik grammatik xususiyatlarni hisobga olishga yordam beradi, lekin har doim mukammal bo‘lmasligi mumkin. Chuqur o‘rganish va mashina o‘rganish metodlari esa tilning morfologiyasini aniq tahlil qilish imkonini beradi, ammo katta hisoblash resurslari talab etadi. Gibrid yondashuvlar bu metodlarni birlashtirib, yuqori samaradorlikni ta'minlashga yordam beradi.

Kalit so‘zlar : NLP, Lemmatizatsiya, Morfologik tahlil, so‘z shakli, spacy, stemming, model.

Annotation: Identifying the root form of words (stemming) plays a crucial role in natural language processing (NLP) as it forms the foundation of morphological analysis. By determining word forms, grammatical and semantic features are identified, which helps in understanding the language and developing relevant technologies. This process is essential in NLP applications such as text classification, translation, and communication systems. Various methods are available to identify word forms, including rule-based approaches, morphological modules, lemmatization, stemming, machine learning, and deep learning techniques. Each approach has its own advantages and drawbacks. For example, rule-based methods are simple and effective but may not fully capture complex grammatical structures. Morphological modules provide high accuracy but require large datasets. Lemmatization and stemming methods help account for subtle grammatical features in the language, though they are not always perfect. Deep learning and machine learning methods allow for more accurate analysis of language morphology but require significant computational

Keywords: NLP, Lemmatization, Morphological analysis, word form, spacy, stemming, model.

Аннотация: Определение корневой формы слов (стемминг) играет ключевую роль в обработке естественного языка (NLP), поскольку является основой морфологического анализа. Определяя формы слов, выявляются грамматические и семантические особенности, что помогает в понимании языка и разработке соответствующих технологий. Этот процесс необходим в приложениях NLP, таких как классификация текста, перевод и системы общения. Существуют различные методы для определения форм слов, включая основанные на правилах подходы, морфологические модули, лемматизацию, стемминг, методы машинного обучения и глубокого обучения. Каждый метод имеет свои преимущества и недостатки. Например, методы, основанные на правилах, просты и эффективны, но не всегда могут полностью охватить сложные грамматические структуры. Морфологические модули обеспечивают высокую точность, но требуют больших объемов данных. Лемматизация и стемминг помогают учитывать тонкие грамматические особенности языка, хотя не всегда идеальны. Методы глубокого обучения и машинного обучения позволяют более точно анализировать морфологию языка, но требуют значительных вычислительных ресурсов. Гибридные подходы комбинируют эти методы для обеспечения большей эффективности и лучшей производительности.

Ключевые слова: NLP, лемматизация, морфологический анализ, словоформа, пространственность, стемминг, модель.

Tabiiy tilni qayta ishlash (NLP) sohasida soʻz shakllarining asosini aniqlash, yaʼni ildiz shaklini topish, juda muhim ahamiyatga ega. Soʻz shakllari morfologik tahlilning asosiy obʼekti boʻlib, bu jarayon tilni toʻgʻri tushunish va unga asoslangan texnologiyalarni ishlab chiqish uchun zarurdir. Soʻz shakllarini aniqlash orqali soʻzning oʻziga xos grammatik va semantik xususiyatlari aniqlanadi. Bu jarayonning muvaffaqiyatli amalga oshirilishi tilni qayta ishlashning bir qancha bosqichlarida, xususan, maʼlumotni tasniflash, tarjima qilish, muloqot tizimlari yaratish va boshqa koʻplab NLP dasturlarida muhim rol oʻynaydi. Soʻz shakllarining ildizini aniqlashga yordam beruvchi algoritmlar turli tillarda keng qoʻllaniladi, ammo har bir tilning oʻziga xos grammatik tizimi bu jarayonni murakkablashtiradi.

Yondashuvlar va Metodlar. Soʻz shakllarining asosini aniqlash jarayoni turli metodlar va yondashuvlar asosida amalga oshiriladi. Ushbu metodlar nafaqat soʻz shakllarini aniqlashning aniq usullarini taklif etadi, balki ularni amalga oshirishda yuzaga keladigan qiyinchiliklarni ham yengishga harakat qiladi. Morfologik tahlilning bir nechta asosiy yondashuvlari mavjud, ularning har biri oʻzining afzalliklari va kamchiliklariga ega. Ushbu metodlar orasida qoidalarga asoslangan yondashuvlar, morfologik modullar, lemmatizatsiya va stemming, mashina oʻrganish va chuqur oʻrganish usullari va statistik modellar kabi yondashuvlar mavjud. Har bir yondashuvning oʻziga xos ijobiy va salbiy tomonlari mavjud boʻlib, ularning samaradorligi tildan tildan farq qiladi.

Qoidalarga Asoslangan Yondashuvlar: Qoidalarga asoslangan yondashuv tilning grammatik xususiyatlariga tayangan holda soʻz shakllarini aniqlashni oʻz ichiga oladi. Bu metod tilning morfologik tizimining aniq va qatʼiy qoidalariga asoslanadi. Masalan, turkcha tilida qoʻshimchalar va soʻzlarning oʻzgarishi aniq qoidalarga binoan amalga oshiriladi. Shu sababli, qoidalarga asoslangan yondashuv oʻzining oddiyligi va shaffofligi bilan ajralib turadi. Bu yondashuvda soʻzning oxiridagi qoʻshimchalar va ildiz shakli aniqlanadi. Biroq, qoidalarga asoslangan yondashuv tillarda turli istisnolar va murakkabliklarni inobatga olishda qiyinchiliklarga olib kelishi mumkin. Bu metod koʻpincha tilning murakkab grammatik strukturalarini toʻliq aks ettira olmaydi. Qoidalarga asoslangan yondashuvning afzalliklari orasida uning soddaligi va aniq natijalar beradi degan fikr mavjud. Masalan, maʼlum bir tilning grammatikasini yaxshi bilgan holda, soʻz shakllarini aniqlash jarayoni tez va aniq amalga oshiriladi. Ammo, bu metodning kamchiligi shundaki, murakkab grammatik strukturalar, istisnolar va tilning dinamik oʻzgarishlarini hisobga olishda qiyinchiliklar yuzaga kelishi mumkin. Shu sababli, baʼzi hollarda qoidalarga asoslangan metodlar tilni toʻliq tahlil qilishda yetarli boʻlmaydi.

Soʻz shakllarini aniqlash uchun zamonaviy morfologik modullar, masalan, **Mystem**, **Pymorphy2**, **spaCy** kabi vositalar keng qoʻllaniladi. Ushbu modullar morfologik tahlilni amalga oshirishda qoidalarga asoslangan metodlarni va statistik usullarni birlashtiradi. Morfologik modullar yordamida soʻzlarning ildizlari, qoʻshimchalari, va grammatik xususiyatlari aniqlanadi. Bu usulda, masalan, “yurgan” soʻzi feʼlning yurgan shakli boʻlib, lemmatizatsiya orqali “yurmoq” lemmasiga qaytariladi. Morfologik modullar tillarning turli variantlari va istisnolarini oʻz ichiga olgan holda yuqori darajada aniqlikni taʼminlaydi.

Morfologik modullarni qoʻllashning afzalliklari orasida yuqori aniqlik va samaradorlikni taʼminlashi, shuningdek, tilning grammatik qoidalarini va tuzilmalarini yaxshiroq oʻrganishga imkon berish kabilar mavjud. Shu bilan birga, morfologik modullar koʻplab tillarda ishlashga moslashgan va bu usulni xalqaro miqyosda qoʻllash mumkin. Biroq, morfologik modullarni toʻgʻri ishlatish uchun maʼlumotlarning katta hajmda boʻlishi kerakligi va baʼzi murakkab tillarda ularning samaradorligi pasayishi mumkin. Bu tizimlar toʻliq oʻrganish uchun koʻproq vaqt va resurslarni talab qiladi.

Lemmatizatsiya va **stemming** metodlari morfologik tahlilning ikki asosiy usuli boʻlib, soʻz shakllarining ildizlarini aniqlashda keng qoʻllaniladi. Lemmatizatsiya usulida, soʻzlar grammatik jihatdan tahlil qilinib, ularning lugʻat shaklini aniqlashga harakat qilinadi. Masalan, “yurgan”, “yurmoq” kabi soʻzlar barchasi “yurmoq” lemmasiga qaytariladi. Stemming usulida esa soʻzning ildiz shakli aniqroq va tezroq aniqlanadi, masalan, “yurgan” soʻzi ildiz shakliga “yur” qaytariladi. Lemmatizatsiya metodining afzalliklari orasida uning soʻzlarning grammatik xususiyatlarini inobatga olishi va natijalarning yuqori aniqlikda boʻlishi bor. Bu metod tilning nozik grammatik strukturalarini aks ettiradi. Biroq, lemmatizatsiya baʼzan tildagi barcha nuanslarni aks ettira olmaydi, ayniqsa morfologik jihatdan murakkab tillarda. Stemming esa tez va samarali ishlaydi, lekin semantik tafovutlar va noaniqliklar yuzaga kelishi mumkin. Shu sababli, lemmatizatsiya va stemming metodlarini qoʻllashda, tildan foydalanish va murakkabliklarni inobatga olish zarur.

Mashina Oʻrganish.

Mashina oʻrganish va chuqur oʻrganish metodlari soʻz shakllarining asosini aniqlash jarayonida samarali yondashuvlarni taklif etadi. Mashina oʻrganish usullaridan, masalan, **supervised learning** yordamida tilning morfologiyasini aniqlash mumkin. Ushbu usulda, maʼlumotlar toʻplami asosida model oʻrgatiladi, va model oʻzining oʻrganilgan bilimlari asosida yangi soʻz shakllarini aniqlaydi. Chuqur oʻrganish metodlari, masalan, **LSTM**, **GRU**, va **Transformer** kabi neyron tarmoqlari yordamida morfologik tahlilni amalga oshiradi va tilda yuzaga keladigan noaniqliklarni yaxshiroq hal qilishga yordam beradi. Chuqur oʻrganish metodlarining afzalliklari orasida ularning tilni chuqur oʻrganish imkoniyatini yaratishi va maʼlumotlar asosida soʻz shakllarining aniq va toʻliq tahlilini amalga oshirishi mavjud. Biroq, bu metodlar yuqori hisoblash quvvatlari va katta miqdordagi annotatsiyalangan maʼlumotlarni talab qiladi, bu esa baʼzan resurslar nuqtai nazaridan qiyinchiliklarga olib keladi. Shu sababli, chuqur oʻrganish metodlarini qoʻllashda zarur resurslar va maʼlumotlarga ehtiyoj bor.

Statistik modellar, masalan, **Hidden Markov Model (HMM)** va boshqa ehtimollarga asoslangan modellar soʻz shakllarini aniqlashda samarali ishlaydi. Bu modellarda, soʻz shakllari oʻrtasidagi bogʻlanishlar va ehtimollar tahlil qilinadi, va bu orqali soʻzlarning ildiz shakllari aniqlanadi. Statistik modellarning afzalliklari orasida ular katta maʼlumotlar asosida ishlash va ehtimollarni tahlil qilish orqali yuqori samaradorlikka erishadi. Ammo, baʼzan ehtimollarni aniq tahlil qilishda baʼzi noaniqliklar yuzaga kelishi mumkin.

Gibrid yondashuvlar yuqoridagi metodlarni birlashtirib, soʻz shakllarini aniqlashning samaradorligini oshirishga yordam beradi. Masalan, qoidalarga asoslangan yondashuvni mashina oʻrganish metodlari bilan birlashtirish orqali murakkab tillarda yuqori aniqlikni taʼminlash mumkin. Bu yondashuvlar nafaqat soʻz shakllarini aniqlashda, balki tilning grammatik va semantik xususiyatlarini toʻliq aks ettirishda ham foydalidir.

Xulosa

Soʻz shakllarining asosini aniqlash algoritmlari tabiiy tilni qayta ishlashning eng muhim qismlaridan biridir. Ushbu algoritmlar yordamida soʻz shakllarining ildizlari, qoʻshimchalari va grammatik tuzilmalari aniqlanadi. Morfologik tahlilning samarali amalga oshirilishi uchun turli metodlarning birlashtirilgan yondashuvlari kerak boʻladi. Yangi yondashuvlar va metodlarni ishlab chiqish tilni qayta ishlashning yanada samarali va aniqlikni oshiruvchi imkoniyatlarini taqdim etadi.

Foydalanilgan adabiyotlar:

1. Prutskow A.V. Metody poiska reshenij v lingvisitcheskix avtomatizirovannyh
2. obuchajuschih sistemah // Nauchno-tehnicheskaja informacija. Seriya 2: Informatsionnye protsessy i sistemy. – 2006. – N 4. – PP. 15-18.
3. Sokolova, M., & Inkpen, D. (2009). *A Survey of Lexical Semantics in Natural Language Processing*. Journal of Computational Linguistics, 35(3), 1-25.
4. Tufis, D., & Dinu, L. (2012). *Morphological Analysis of Words in Natural Language Processing: A Survey*. Journal of Language Resources and Evaluation, 46(4), 273-287.
5. Aripov M., Fayzullaeva Z.I., Karimov N.N. Annotation of texts based on semantic analysis // AIP Conference Proceedings. – 2024. – Vol. 3244. – P.1030049-1 – 03-0049-8, doi: 10.1063/5.0241988.
6. Fayzullayeva, Z., Olimova, M., Karimov, N. (2024). Abstraktiv annotatsiyalash yordamida matnlarni annotatsiyalovchi mastat tizimini yaratish. *Digital transformation and artificial intelligence*, 2(3), 17-23.
7. Fayzullayeva, Z., Karimov, N., & Abdusattarov, A. (2024). Matndan sunʼiy intellekt yordamida Ekstrativ xulosa olish: Matndan sunʼiy intellekt yordamida Ekstrativ xulosa olish. *Modern Problems and Prospects of Applied Mathematics*, 1(01).
8. Inatillayevna, F. Z., & Nosirjon oʻgʻli, K. N. (2024). NLPda matnlarni umumiyashtirish va leksik tahlil. «Contemporary Technologies of Computational Linguistics», 2(22.04),
9. Kondrak, G. (2001). *Text Classification Using WordNet-Based Concepts*. In *Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Text Processing* (pp. 1-15). Springer

MASHINALI O'RGATISH USULLARI YORDAMIDA MATNNI SENTIMENT TAHLIL QILISH

Navruzov Erkin Ravshanovich

O'zbekiston Milliy universiteti
Amaliy matematika va intellektual texnologiyalar
fakulteti katta o'qituvchisi

Jo'raqulova Madina Subhonovna

O'zbekiston Milliy universiteti
Kompyuter lingvistikasi va amaliy tilshunoslik
kafedrası o'qituvchisi

Annotatsiya. Matn tonalligini aniqlash (sentiment analysis) tabiiy tilni qayta ishlash (NLP) sohasidagi dolzarb masalalardan biridir. Ushbu metod yordamida matnning hissiy holati, ya'ni uning ijobiy, salbiy yoki neytral ekanligi aniqlanadi. Ushbu maqolada matn tonalligini aniqlash metodlari va ularning mashinali o'rgatish va chuqur o'rgatish asosidagi yondashuvlari tahlil qilinadi. Dastlab, matn tonalligini aniqlashning nazariy asoslari va uning dasturiy vositalar yordamida qo'llanilishi ko'rib chiqiladi.

Shuningdek, ushbu maqolada mashinali o'rgatish (machine learning) usullari, jumladan, Naive Bayes, Support Vector Machine (SVM) va Logistic Regression kabi klassik algoritmlar yordamida matn tonalligini aniqlash usullari tavsiflanadi. Shuningdek, chuqur o'rgatish (deep learning) metodlari, xususan, sun'iy neyron tarmoqlari va zamonaviy transformer modellarining imkoniyatlari tahlil etiladi. Ushbu metodlar matnni kontekstual tahlil qilishda va murakkab grammatik tuzilmalarni aniqlashda qanday samarali ishlashini ko'rsatadi.

Kalit so'zlar: Matn tonalligi, sentiment analysis, mashinali o'rgatish, chuqur o'rgatish, sun'iy neyron tarmoqlari, Naive Bayes, Support Vector Machine, Logistic Regression, Tabiiy tilni qayta ishlash (NLP).

Аннотация. Анализ тональности текста является одной из ключевых задач в области обработки естественного языка (NLP). Этот метод позволяет определить эмоциональное состояние текста, будь то положительное, отрицательное или нейтральное. В данной работе анализируются методы определения тональности текста и их подходы, основанные на машинном обучении и глубоком обучении. Вначале рассматриваются теоретические основы анализа тональности и его применение с использованием программных инструментов.

В работе описаны методы машинного обучения, такие как Naive Bayes, Support Vector Machine (SVM) и Logistic Regression, для классификации тональности текста. Также рассматриваются методы глубокого обучения, особенно искусственные нейронные сети и современные трансформерные модели, которые позволяют учитывать контекст и распознавать сложные грамматические структуры.

Ключевые слова: Тональность текста, анализ тональности, машинное обучение, глубокое обучение, искусственные нейронные сети, Naive Bayes, Support Vector Machine, Logistic Regression, обработка естественного языка (NLP).

Abstract. Sentiment analysis is one of the crucial tasks in the field of Natural Language Processing (NLP). This method determines the emotional state of the text, whether it is positive, negative, or neutral. This thesis analyzes the methods of sentiment analysis and their approaches based on machine learning and deep learning. First, the theoretical foundations of sentiment analysis and its application using software tools are discussed.

In the thesis, machine learning methods such as Naive Bayes, Support Vector Machine (SVM), and Logistic Regression are described for sentiment classification. Additionally, deep learning methods, particularly artificial neural networks and modern transformer models, are analyzed for their ability to contextualize and recognize complex grammatical structures.

Keywords: Text sentiment, sentiment analysis, machine learning, deep learning, artificial neural networks, Naive Bayes, Support Vector Machine, Logistic Regression, natural language processing (NLP).

Kirish. So‘nggi yillarda internet tarmog‘ida foydalanuvchilar tomonidan yaratilayotgan matnli ma‘lumotlar soni keskin ortdi. Ijtimoiy tarmoqlardagi fikrlar, mahsulotlar haqidagi izohlar, yangiliklarga bildirilgan munosabatlar va sharhlar bugungi kunda turli sohalarida qaror qabul qilish jarayoniga ta‘sir ko‘rsatmoqda. Bu matnlar tarkibidagi axborotlarni tahlil qilish va ulardan foydali xulosalar chiqarish ehtiyoji tabiiy ravishda matn tonalligini aniqlash (sentiment analysis) sohasining rivojlanishiga olib keldi.

Sentiment tahlil – bu matnning ijobiy, salbiy yoki neytral ekanligini aniqlashga qaratilgan tabiiy tilni qayta ishlash (Natural Language Processing, NLP) yo‘nalishidagi muhim masalalardan biridir. Ushbu yondashuv orqali kompaniyalar mijozlarning mahsulot yoki xizmatlarga bo‘lgan munosabatini o‘rgatish, siyosiy fikrlarni tahlil qilish yoki jamoatchilik kayfiyatini o‘lchash imkoniyatiga ega bo‘ladilar.

Avvalgi yillarda sentiment tahlilni amalga oshirishda lug‘aviy yondashuvlar (ya‘ni ijobiy yoki salbiy so‘zlar ro‘yxatiga asoslangan baholash) keng qo‘llanilgan bo‘lsa, bugungi kunda bu masalaga mashinali o‘rgatish (machine learning) va chuqur o‘rgatish (deep learning) metodlari asosida yondashilmoqda. Ushbu metodlar ko‘plab murakkab til hodisalarini hisobga olish orqali tonallikni yanada aniqroq aniqlashga imkon beradi.

Mazkur maqolada matn tonalligini aniqlash masalasiga zamonaviy mashinaviy o‘rgatish algoritmlari orqali yondashiladi, ularning turlari, ustunliklari hamda o‘zbek tilidagi amaliy qo‘llanish imkoniyatlari tahlil qilinadi.

Matn tonalligini aniqlash (sentiment analysis) — bu matn mazmunidagi hissiy munosabatni aniqlovchi tabiiy tilni qayta ishlashning tarmoqlaridan biridir. Ushbu jarayonda matn tahlil qilinib, undagi kayfiyat yoki munosabatning ijobiy, salbiy yoki neytral ekani belgilanadi. Masalan, “Ushbu mahsulot juda zo‘r!” kabi jumla ijobiy tonallikda, “Xizmat sifati yomon” degani esa salbiy tonallikda yozilgan deb baholanadi.

Tonallikni aniqlash uchun dastlab korpus (ya‘ni, belgilangan — *annotatsiya qilingan* matnlar to‘plami) talab etiladi. Annotatsiya qilish jarayoni odatda mutaxassislar tomonidan matnlarning hissiy yo‘nalishini belgilash orqali amalga oshiriladi. Olingan korpus asosida algoritmlar o‘rgatiladi va yangi, noma‘lum matnlarning tonalligi bashorat qilinadi.

Mashinali o‘rgatish asoslari: Matn tonalligini aniqlashda mashinali o‘rgatish algoritmlari keng qo‘llaniladi. Ushbu yondashuvda har bir matn sonli shaklga keltiriladi va keyin model uni o‘rganib chiqadi. Mashhur algoritmlar quyidagilardan iborat:

- **Naive Bayes** – matnlarni ehtimollik asosida klassifikatsiya qiluvchi oddiy, lekin samarali usul;
- **Support Vector Machine (SVM)** – murakkab va chiziqli bo‘lmagan chegaralarni topishda kuchli klassifikator;
- **Logistic Regression** – tonallikni ehtimollik asosida aniqlash uchun ishlatiladi va ko‘plab NLP masalalarida muvaffaqiyatli qo‘llaniladi.

Bu algoritmlarni qo‘llashdan oldin matnni raqamli ko‘rinishga keltirish kerak bo‘ladi. Bunda quyidagi metodlar qo‘llaniladi:

- **Bag of Words (So‘zlar xaltasi modeli)** – matndagi har bir so‘z alohida xususiyat sifatida olinadi;
- **TF-IDF (Term Frequency – Inverse Document Frequency)** – so‘zning muhimligini hisobga oluvchi vaznlashtirish usuli;

• **Word2Vec** – har bir soʻzni semantik maʼnosi bilan vektor koʻrinishda ifodalovchi neyron tarmoq modeli.

Modelni baholash uchun esa quyidagi metrikalar qoʻllaniladi:

- **Accuracy (aniqlik)** – umumiy toʻgʻri bashoratlar ulushi;
- **Precision va Recall** – ayniqsa, ijobiy/salbiy klasslar uchun muhim;
- **F1-score** – precision va recall oʻrtasidagi muvozanat.

Ushbu yondashuvlar oʻzbek tilidagi matnlar bilan ham ishlashi mumkin, biroq korpus yetishmasligi, til qoidalarining oʻziga xosligi baʼzi murakkabliklarni yuzaga keltiradi.

Chuqur oʻrgatish va zamonaviy yondashuvlar:

Klassik mashinali oʻrgatish usullari baʼzi hollarda samarali natija bersa-da, matn mazmunini chuqur tushunish, kontekstni aniqlash va murakkab grammatik tuzilmalarga moslashishda cheklovlariga ega. Shu sababli, soʻnggi yillarda **chuqur oʻrgatish (deep learning)** asosidagi yondashuvlar matn tonalligini aniqlashda keng qoʻllanila boshlandi.

Chuqur oʻrgatish yondashuvi, asosan, **sunʼiy neyron tarmoqlarga** tayanadi. Ushbu usullar matnni ketma-ketlik sifatida qabul qilib, ularni kontekstual tarzda tahlil qiladi. Quyidagi arxitekturalar bu yoʻnalishda eng samaralilardan hisoblanadi:

• **LSTM (Long Short-Term Memory)** – vaqt ketma-ketligidagi uzoq masofali bogʻlanishlarni hisobga oluvchi rekurrent neyron tarmoq turi;

• **BiLSTM (Bidirectional LSTM)** – matnni ikki yoʻnalishda oʻqish orqali har bir soʻzga toʻliq kontekstni beradi;

• **GRU (Gated Recurrent Unit)** – LSTMga oʻxshash, ammo soddalashtirilgan versiya, tezroq ishlaydi.

Ushbu arxitekturalar kontekstni chuqur tushunadi va jumla ohangini aniqlashda yuqori aniqlik koʻrsatadi. Biroq, eng katta yutuq **Transformer** arxitekturasining paydo boʻlishi bilan boshlandi.

Transformer modellar: BERT va uning variantlari

2018-yilda Google tomonidan ishlab chiqilgan **BERT (Bidirectional Encoder Representations from Transformers)** modeli NLP sohasida inqilob yasadi. U matnni ikki tomondan oʻqiydi (chapdan va oʻngdan), yaʼni soʻzlarning kontekstual bogʻlanishini chuqur anglaydi. BERT modeli:

- har bir soʻzning kontekstual vektor ifodasini yaratadi;
- tonallik, tasniflash, savol-javob, tarjima kabi topshiriqlarda yuqori aniqlikka ega;
- oldindan oʻqitilgan (pretrained) model boʻlib, uni yangi maʼlumotlarga **moslashtirish (fine-tuning)** oson.

BERT modeli asosida boshqa kuchli modellar ham ishlab chiqilgan:

• **RoBERTa** – BERTʼning yaxshilangan versiyasi, koʻproq maʼlumot va kuchli oʻqitish strategiyasiga ega;

• **DistilBERT** – yengil va tezkor model;

• **mBERT (multilingual BERT)** – koʻp tili matnlar bilan ishlay oladigan versiya;

• **XLM-RoBERTa** – 100+ tilni qoʻllab-quvvatlovchi, BERTʼga nisbatan ham aniqroq model.

Ayni paytda, **oʻzbek tili uchun maxsus moslashtirilgan BERT modellari** hali kam, ammo **mBERT** yoki **XLM-R** yordamida oʻzbek tilidagi matnlarning tonalligini aniqlash mumkin. Bunda maxsus oʻzbek tili korpuslari kerak boʻladi – bu masala esa tadqiqotning alohida dolzarb yoʻnalishlaridan biridir.

Zamonaviy model yordamida tonallikni aniqlashda, quyidagi bosqichlar bajariladi:

1. Matn tozalab olinadi va modelga yuboriladi;
2. Model har bir soʻzga kontekstual vektor beradi;
3. Ehtimolliklar asosida ijobiy/salbiy/neytral baho chiqariladi.

Bu jarayonni Python tilidagi kutubxonalar, masalan, **HuggingFace Transformers**, **TensorFlow**, yoki **PyTorch** orqali amalga oshirish mumkin.

Matn tonalligini aniqlash amaliyotda ko‘plab dasturiy vositalar yordamida amalga oshiriladi. Python dasturlash tili bu borada eng keng qo‘llaniladigan vositalardan biridir. U orqali turli mashinali o‘rgatish kutubxonalari, neyron tarmoqlar va tayyor transformer modellaridan foydalanish mumkin. Quyida asosiy kutubxonalar va qisqacha amaliy yondashuv keltiriladi:

Python kutubxonalari:

- **scikit-learn** – klassik algoritmlar (Naive Bayes, SVM, Logistic Regression) uchun;
- **nlTK va spaCy** – matnni tozalash, tokenizatsiya va sintaktik tahlil uchun;
- **TensorFlow va Keras** – neyron tarmoqlarni yaratish uchun;
- **transformers (HuggingFace)** – BERT va boshqa tayyor modeldan foydalanish uchun.

Oddiy misol: BERT orqali tonallik aniqlash

```
from transformers import pipeline
model = pipeline("sentiment-analysis", model="nlptown/bert-base-multilingual-uncased-sentiment")
text = "Bu mahsulot menga juda yoqdi, sifati a'lo darajada!"
result = model(text)
print(result)
```

Natija:

```
[{'label': '5 stars', 'score': 0.87}]
```

Yuqoridagi kodda **nlptown/bert-base-multilingual-uncased-sentiment modeli** ingliz, fransuz, nemis, ispan va boshqa tillar bilan bir qatorda o‘zbek tilidagi oddiy gaplarni ham qisman tahlil qila oladi. Bu model yordamida o‘zbek tilida yozilgan izohlar mazmunini aniqlash mumkin, biroq natijani yanada aniqlashtirish uchun o‘zbek tilida moslashtirilgan (fine-tuned) model kerak bo‘ladi.

Amaliy tajriba shuni ko‘rsatadiki, o‘zbek tilidagi tonallikni aniq baholash uchun hali yetarli annotatsiyalangan korpuslar va tilga moslashtirilgan modellar mavjud emas. Bu esa o‘zbek tili uchun maxsus NLP tadqiqotlar olib borish zarurligini ko‘rsatadi. Shu bilan birga, mavjud ko‘p tilli modellar orqali boshlang‘ich darajadagi sentiment analiz tizimlarini yaratish mumkin.

Xulosa. Matn tonalligini aniqlash (sentiment analysis) so‘zi bugungi kunda sun‘iy intellekt va tabiiy tilni qayta ishlash (NLP) sohalarida dolzarb masala hisoblanadi. Ushbu maqolada matn tonalligini aniqlash metodlarini mashinali o‘rgatish usullari yordamida o‘rgatish va zamonaviy chuqur o‘rgatish modellarining imkoniyatlari tahlil qilindi.

Klassik mashinali o‘rgatish algoritmlari (Naive Bayes, SVM, Logistic Regression) matn tonalligini aniqlashda ishlatilishi mumkin, ammo bu metodlarning aniqligi va samaradorligi ko‘proq kichik va o‘rtacha hajmdagi korpuslar bilan cheklangan. Ular matnlarni sonli shaklda ifodalashda samarali bo‘lsa-da, kontekstual bog‘lanishlarni to‘liq anglay olmaydi.

Bugungi kunda **chuqur o‘rgatish modellari** – ayniqsa, **Transformer arxitekturasini** asosidagi modellar (BERT, RoBERTa, XLM-R) yordamida matn tonalligini aniqlash ancha yuqori aniqlikda amalga oshiriladi. Bu modellar tilning chuqur tuzilishini, jumlar orasidagi kontekstual bog‘lanishlarni hisobga olish imkoniyatini beradi, bu esa ular uchun ijobiy va salbiy tonallikni aniqlashni ancha osonlashtiradi.

Shuningdek, **o‘zbek tilida** maxsus ishlangan model va korpuslar kamligi katta muammo hisoblanadi. O‘zbek tilidagi annotatsiyalangan matnlar yetishmasligi tonallik tahlili bo‘yicha ishlanmalarning rivojlanishini cheklaydi. Shu sababli, o‘zbek tilida **maxsus NLP modellarini ishlab chiqish va korpuslarni kengaytirish** juda muhimdir.

Foydalanilgan adabiyotlar:

1. Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers.
2. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.
3. Vinyals, O., & Le, Q. V. (2015). “A Neural Network for Factored Translation Models.” *Proceedings of the 32nd International Conference on Machine Learning*.
4. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). “Distributed Representations of Words and Phrases and their Compositionality.” *Proceedings of NeurIPS 26*.
5. Scikit-learn Documentation. (2023). “Machine Learning in Python.” <https://scikit-learn.org/stable/index.html>.
6. Hirschberg, J., & Manning, C. D. (2015). “Natural Language Processing.” *Communications of the ACM*, 58(6), 43–50.
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. *Proceedings of NeurIPS 30*.

N-GRAM INDEKSLARIDAN FOYDALANADIGAN IMLO TUZATISH ALGORITMI

Ahmedova Maftuna

Alisher Navoiy nomidagi

Toshkent davlat o'zbek tili va adabiyoti universiteti

Kompyuter lingvistikasi mutaxassisligi magistranti

E-mail: MaftunaAhmedova1997@gmail.com

Annotatsiya. Imlo xatolarini avtomatik to'g'rilash funksiyasi tizimlarida keng qo'llaniladi. Uning asosiy vazifasi – matndagi imloviy xatolarini aniqlash va ularni tuzatish orqali foydalanuvchilarning vaqtini tejashdir. Lekin lug'atda millionlab qatorlar mavjud bo'lganda, eng yaqin mos keluvchi qatorni izlab topish ko'p vaqt talab qilishi mumkin. Avtomatik tuzatish funksiyasining samaradorligini oshirish uchun turli yechimlar taklif etilgan. Ushbu yondashuvlarning barchasi muammolarni hal qilish uchun tegishli ma'lumotlar tuzilmasi va ba'zi gipotezalar asosida ishlaydi. N-gram til modelidan foydalanib, lug'atda noaniq qatorlarni filtrlagan holda ishlashga imkon beruvchi yangi gibril algoritmnining imlo tuzatishdagi samaradorligini ko'rib chiqamiz. Ushbu yondashuv, lug'atda aniqlikni ta'minlab, xatolarni tuzatishda yuqori samaradorlikni taqdim etadi. Tajribalar shuni ko'rsatmoqdaki, taklif qilingan algoritim tezkor va aniq natijalar olishda juda samarali hisoblanadi.

Kalit so'zlar: tahrir masofasi; n-gram; trigram; satr qidirish; Levenshtein masofasi.

Annotation. The automatic spelling correction feature is widely used in various systems. Its main task is to identify and correct spelling errors in text, thereby saving users' time. However, when the dictionary contains millions of entries, finding the most relevant match can be time-consuming. To improve the efficiency of automatic correction, various solutions have been proposed. All of these approaches operate based on appropriate data structures and certain hypotheses to solve the problem. In this study, we explore the effectiveness of a new hybrid algorithm for spelling correction that leverages an N-gram language model to filter uncertain entries in the dictionary. This approach ensures accuracy in the dictionary while providing high efficiency in correcting errors. Experiments show that the proposed algorithm is highly effective in delivering fast and accurate results.

Keywords: edit distance; n-gram; trigram; string search; Levenshtein distance.

Аннотация. Функция автоматической коррекции орфографии широко используется в различных системах. Её основная задача — выявление и исправление орфографических ошибок в тексте, что позволяет сэкономить время пользователей. Однако при наличии миллионов записей в словаре поиск наиболее подходящего совпадения может занимать много времени. Для повышения эффективности автоматической коррекции были предложены различные решения. Все эти подходы основаны на использовании соответствующих структур данных и определённых гипотез для решения проблемы. В данном исследовании рассматривается эффективность нового гибридного алгоритма коррекции орфографии, использующего языковую модель N-грамм для фильтрации неопределённых записей в словаре. Такой подход обеспечивает точность в словаре и высокую эффективность при исправлении ошибок. Эксперименты показывают, что предлагаемый алгоритм весьма эффективен для получения быстрых и точных результатов.

Ключевые слова: расстояние редактирования; n-грамма; триграмма; поиск строк; расстояние Левенштейна.

Kirish

Hozirgi kunda imloni avtomatik tuzatish tizimlari ko'plab sohalarda keng qo'llanilmoqda. Imloni avtomatik tuzatuvchi tizim foydalanuvchi tomonidan kiritilgan noto'g'ri so'zni aniqlab, uni

avtomatik ravishda to'g'irlashni ta'minlaydi. Masalan, Google qidiruv tizimida foydalanuvchi qidiruv tizimiga noto'g'ri matn kiritganida, tizim ushbu matnni to'g'irlab, qidiruv natijalarini to'g'irlangan matn asosida taqdim etadi.

Avtomatik tuzatish funksiyasining asosiy tamoyili kiritilgan so'z yoki matnga eng yaqin bo'lgan so'z yoki matnni aniqlashga asoslanadi. Ikki matn orasidagi minimal farqlar soni yoki bir matnni boshqasiga aylantirish uchun zarur bo'lgan jarayon tahrir masofasi (edit distance) deb nomlanadi. Ushbu masofa ma'lumotlar bazasidagi har bir so'z va foydalanuvchi tomonidan kiritilgan matn o'rtasida hisoblanadi. Minimal tahrir masofasiga ega bo'lgan so'z qidiruv natijasi sifatida tanlanadi.

Ko'plab ilovalarda avtomatik tuzatish funksiyasi muvaffaqiyatli joriy qilingan. Biroq, ulkan hajmdagi lug'atlar bilan ishlashda, har bir matn juftligi uchun tahrir masofasini hisoblash jarayoni sezilarli vaqt talab qiladi. Bu ayniqsa, bulutli xizmatlarda, xususan, Google qidiruv tizimi kabi platformalarda, resurslardan samarasiz foydalanish muammosini keltirib chiqaradi.

Bulutli xizmatlar odatda mashina darajasidagi cheklangan API vositalaridan foydalanadi, natijada barcha hisob-kitoblar yuqori darajadagi API orqali amalga oshadi. Bunday yondashuv lug'atdagi barcha so'zlar uchun tahrir masofasini hisoblash jarayonini sekinlashtiradi va bu real vaqt rejimida ishlash imkoniyatini cheklaydi. Natijada, qidiruv tizimlarida tahlil qilish va foydalanuvchilarga aniq natijalarni taqdim etish qiyinlashadi. Mazkur muammoni bartaraf etish uchun yangi algoritmlar va yondashuvlardan foydalanish zarur. Ushbu yondashuvlar tahrir masofasini hisoblashni tezlashtirishga va samaradorlikni oshirishga qaratilgan. N-gramga asoslangan indekslash kabi texnologiyalar orqali resurslardan samarali foydalanish va noaniq ma'lumotlarni filtrlash imkoniyati ta'minlanadi. Tajribalar bu kabi usullar muvaffaqiyatli natijalar berishini ko'rsatmoqda.

Mazkur muammoni hal etish bo'yicha chet el mutaxassislari tomonidan turli yondashuvlar taklif qilingan. Jong Yong Kim va John Shawe-Taylor[1] DNK ketma-ketliklari uchun algoritmlar ishlab chiqqan bo'lsa, Klaus U. Schulz va Stoyan Mihov Levenshtein masofasi algoritmi asosida chekli holatli avtomatlar yordamidagi usul taklif qilganlar[2]. Victoria J. Hodge va Jim Austin foydalanuvchi so'rovlarini tekshirish uchun Hamming masofasi va N-gram algoritmlarini birlashtiruvchi gibrid metodologiyani tavsiya etganlar[3]. Ammo Hamming masofasi algoritmi tahrir masofasini hisoblashda faqat teng uzunlikdagi matnlarni talab qiladi.

Mazkur tadqiqotda Levenshtein masofasi algoritmini takomillashtirishdan ko'ra, muammoni hal qilish uchun muqobil yondashuv taklif etildi. Taklif etilgan yondashuv lug'atdagi barcha so'zlar uchun tahrir masofasini hisoblash o'rniga, faqat ehtimoliy mos keluvchi so'zlar to'plamini oldindan aniqlash va ularni filtrlagan holda hisoblashni amalga oshiradi. Ushbu maqolada biz ushbu yondashuvga asoslangan yangi algoritmlarni taqdim etamiz. Mazkur algoritmlar so'zlarni filtrlash jarayonini amalga oshirib, tahrir masofasini faqat tanlangan so'zlar uchun hisoblaydi, bu esa hisoblash jarayonining samaradorligini oshirish va bajarilish vaqtini sezilarli darajada qisqartirish imkonini beradi. Taklif etilgan algoritmlar yuqori ishlash samaradorligi va qisqartirilgan hisoblash vaqti bilan ajralib turadi.

Tahrir masofasi: ikki satr orasidagi minimal o'zgarishlar soni orqali aniqlanadi. Bu o'zgarishlar quyidagi amallarni o'z ichiga olishi mumkin: kirish (insert), o'chirish (delete), almashtirish (replace) va ba'zan joyini almashtirish (transpose). Qaysi amallar ruxsat etilishi esa tanlangan algoritmlar turiga bog'liq bo'ladi. Tahrir masofasiga asoslangan turli algoritmlar mavjud bo'lib, ular ayniqsa matnni avtomatik tuzatish tizimlarida keng qo'llaniladi. Misol uchun, agar matnda mavjud bo'lmagan biror so'z aniqlansa, ushbu algoritmlar yordamida unga eng yaqin tahrir masofasiga ega bo'lgan so'z topilib, to'g'ri variant sifatida foydalanuvchiga taklif qilinishi mumkin.

Quyida ikki satr o'rtasidagi tahrir masofasini aniqlashda qo'llaniladigan ba'zi algoritmlar keltirilgan:

- Hamming masofasi
- Levenshtein masofasi

- Damerau-Levenshtein masofasi
- Jaro-Winkler masofasi
- Ukkonen algoritmi

Ushbu algoritmlar orasida biz Levenshtein masofasi algoritmini tanlab oldik.

Levenshtein masofasi (LD) ikki satr o'rtasidagi o'xshashlikni aniqlash uchun ishlatiladigan o'lchov hisoblanadi. U ikki satrni bir-biriga aylantirish uchun zarur bo'lgan minimal o'zgartirishlar sonini aniqlaydi. Ushbu o'zgartirishlar bitta belgini kiritish (insertion), o'chirish (deletion) yoki almashtirish (substitution) amallarini o'z ichiga oladi. Ushbu algoritm 1965-yilda rus olimi **Vladimir Levenshtein** tomonidan ishlab chiqilgan va bu algoritm o'zining soddaligi hamda samaradorligi bilan ajralib turadi.

Levenshtein masofasi algoritmidagi kamchiliklar

Levenshtein masofasi algoritmining ishlash vaqt murakkabligi $O(|S_1| \cdot |S_2|)$ ya'ni ikki satr uzunliklari taxminan N bo'lsa, bu $O(N^2)$ ga teng. Algoritmning xotira murakkabligi ham $O(N^2)$ ni tashkil etadi.

Ushbu murakkablik bilan katta hajmli lug'atga ega avtomatik tuzatish funksiyasini ishlab chiqish samarali emas. Chunki Levenshtein masofasi algoritmi $O(N^2)$ murakkablikka ega bo'lib, har bir lug'atdagi so'z uchun bajarilishi kerak. Lug'atdagi barcha so'zlar uchun minimal tahrir masofasiga ega bo'lgan so'z natijaviy so'z sifatida tanlanadi.

Agar lug'atda M ta so'z mavjud bo'lsa, algoritm har bir so'z uchun bajarilib, umumiy vaqt murakkabligi $O(M \cdot N^2)$ ga teng bo'ladi. Bu degani, agar lug'atda 1 million yoki 1 milliard so'z bo'lsa, algoritm mos ravishda 1 million yoki 1 milliard marta bajarilishi kerak. M qiymati oshgan sari, ushbu yechim amaliy jihatdan tobora noqulay bo'lib boradi. Shuning uchun, katta hajmli lug'atlar uchun ushbu algoritmini samarali ishlatish qiyin bo'lib, yangi yondashuv yoki optimallashtirish mexanizmlarini joriy qilish zarur.

N-Gram indekslash

N-gramlar til korpusi yoki deyarli har qanday turdagi ma'lumotlar ketma-ketligidan to'plangan elementlar ketma-ketligidir. Ushbu elementlar qo'llanilish bo'yicha fonemalar, bo'g'inlar, harflar, so'zlar bo'lishi mumkin.

N-gramdagi n hisobga olinadigan bir qator elementlarning o'lchovini, $n=1$ uchun unigram, $n=2$ uchun bigram, $n=3$ uchun trigram va hokazolarni belgilaydi.[5]

N-gram modeli bu ketma-ketlikda keyingi elementni bashorat qilish uchun ishlatiladigan ehtimoliy model hisoblanadi. N-gram modellar statistik tabiiy tilni qayta ishlash va genetik ketma-ketliklarni tahlil qilish sohalarida keng qo'llaniladi.

Taklif etilgan algoritm

Yuqorida keltirilgan muammo shundan iboratki, Levenshtein masofasi algoritmini katta hajmdagi lug'at uchun qo'llashda har bir so'z uchun minimal tahrir masofasini topish juda ko'p vaqt talab qiladi.

Levenshtein masofasi algoritmining vaqt murakkabligi $O(N^2)$ ga teng, shuningdek, lug'atdagi har bir so'z uchun ushbu algoritmini bajarish umumiy murakkablikni $O(M \cdot N^2)$ darajasiga olib keladi. Bu katta hajmdagi lug'atlarda algoritmini qo'llashni amaliy jihatdan samarali bo'lishiga to'sqinlik qiladi. Shu sababli, bunday yondashuvni taklif etdik.

Levenshtein masofasi algoritmi, faqat qidirilayotgan so'zga eng yaqin mos keluvchi so'zlar uchun hisoblash jarayonini amalga oshiradi. Bu esa, ayniqsa katta hajmdagi lug'atlar bilan ishlaganda, tizimning samaradorligini sezilarli darajada oshiradi. Taklif etilgan algoritm quyida **1-jadval**da ko'rsatilgan.

1-jadval. Avtomatik to'g'rilashda vaqtni qisqartirish algoritmi

1-qadam: Qidiruv satri uchun ASCII ro'yxatini yaratish.	$searchAscii = createList(searchTerm) \quad j = 0$
2-qadam: Ma'lumotlar bazasidagi har bir so'z uchun 4-qadamgacha takrorlash.	$for \quad i = 0 \text{ to number of strings in database}$

3-qadam: Ma'lumotlar bazasidagi so'z va qidiruv satri uzunliklari orasidagi farqni hisoblash.	$Diff = abs(len(lstAscii[i]) - len(searchAscii))$
4-qadam: Agar uzunliklar farqi 4 dan kichik bo'lsa, faqat o'sha so'z uchun jarayonni davom ettirish.	<pre> if Diff < threshold_min merge[i] = mergeList(lstAscii[i], searchAscii) if len(merge[i]) < len(searchAscii) + threshold_max stringToPass[j] = i j++ end if end if [Takrorlashning oxiri] </pre>
5-qadam: Minimal qiymatni boshlang'ich qiymat bilan belgilash.	$min = 0$
6-qadam: 8-qadamgacha takrorlash.	$for\ i = 0\ to\ len(stringToPass)$
7-qadam: Levenshtein masofasi algoritmini chaqirish.	<pre> cost[i] = LevenshteinDistance(words[stringToPass[i]], searchTerm) </pre>
8-qadam: Minimal qiymatni tekshirish.	<pre> if cost[min] > cost[i] min = stringToPass[i] end if [Takrorlashning oxiri] </pre>

Eslatma: Bu algoritm Levenshtein masofasi algoritmini faqat qidiruv so'ziga mos keladigan kichikroq tanlangan so'zlar uchun bajarib, umumiy bajarilish vaqtini optimallashtirishga qaratilgan. Bu yondashuv samaradorlikni oshirish va katta lug'atlar bilan ishlashda resurslarni tejash imkonini beradi.

Algoritm

Ushbu algoritm N-gram indeksatsiya konsepsiyasiga asoslanadi. Algoritmning asosiy g'oyasi shundan iboratki, lug'atda saqlangan har bir so'z trigramlarga bo'linadi va qidiruv satri ham trigramlarga ajratiladi. Shundan so'ng, lug'atdagi so'zlar trigramlari va qidiruv satrining trigramlari orasidagi mos kelmaydigan qismlar aniqlanadi.

Mos kelmaydigan trigramlar soniga asoslanib, Levenshtein masofasi hisoblash uchun qaysi so'z tanlanishi belgilanadi. Shuni ta'kidlash lozimki, algoritm ikki trigramni bevosita solishtirmaydi, chunki bu ham $O(N^2)$ murakkablikka olib keladi. Buning o'rniga har bir trigramning harflari ASCII qiymatlarining yig'indisi hisoblanadi. Keyinchalik, solishtirish aynan ushbu ASCII qiymatlar yig'indisi bo'yicha amalga oshiriladi, bu esa $O(N)$ vaqt murakkabligini ta'minlaydi.

Nima uchun trigramlar ishlatiladi? Chunki unigram yoki bigramlardan foydalanilganda, uzunroq satrlar ko'plab kichik subsekvensiyalar yaratadi, bu esa ishlov berish vaqtini oshiradi. Agar trigramlardan uzunroq subsekvensiya yaratilsa, bu natijaning aniqligini pasaytiradi.

Trigramlardan foydalanish natijasida ishlash vaqti jihatidan samarali va kerakli natijaga erishish bo'yicha aniq algoritmgaga ega bo'lamiz. Shu bilan birga, bu yondashuv tezlik va aniqlik o'rtasidagi muvozanatni saqlaydi, ayniqsa katta lug'at hajmiga ega tizimlarda natija yaqqol seziladi.

- **Words[m][n]:** Bu o'zgaruvchi lug'atdagi so'zlar ro'yxatini saqlaydi.
- **searchTerm:** Bu foydalanuvchi kiritgan qidiruv satri.
- **lstAscii[m][n-2]:** Har bir database so'zidagi trigramlarning ASCII qiymatlarini tartiblangan ro'yxatini saqlaydi.
- **searchAscii[]:** Bu qidiruv sadridagi trigramlarning ASCII qiymatlarini tartiblangan ro'yxatidir.

- **merge[m][]**: Ikkita ro'yxatni birlashtirgandan keyin hosil bo'lgan tartiblangan ro'yxat.
- **stringToPass[]**: Levenshtein masofasiga yuboriladigan so'zlar ro'yxati.

1. Jadvalda ko'rsatilgan algoritm databasedagi so'zlar ro'yxati bilan ishlaydi va har bir so'zning trigramlari uchun ASCII qiymatlari yig'indisidan iborat tartiblangan ro'yxatni saqlaydi. Trigram ro'yxatini yaratish databasedagi har bir so'z uchun bir marta bajariladigan hisoblanadi, so'ngra ma'lumotlar bazasida saqlanadi. Algoritmida shuningdek foydalanuvchining kiritgan qidiruv stringi ham mavjud. Algoritm qidiruv sadridagi trigramlarning ASCII qiymatlarini yig'indisini hisoblaydi va ularni tartiblangan ro'yxatga (searchAscii) saqlaydi.

2. Keyinchalik, qidiruv sadrining tartiblangan ro'yxati bilan databasedagi har bir so'zning tartiblangan ro'yxati o'rtasidagi uzunlik farqini solishtiradi. Agar farq **threshold_min** dan kam bo'lsa, faqat bunday ro'yxatlarda merge operatsiyasi amalga oshiriladi. Tadqiqotda **threshold_min** qiymati to'rtga teng, chunki uzunlik farqi uchdan ko'proq bo'lgan stringlar kamdan-kam bo'ladi.

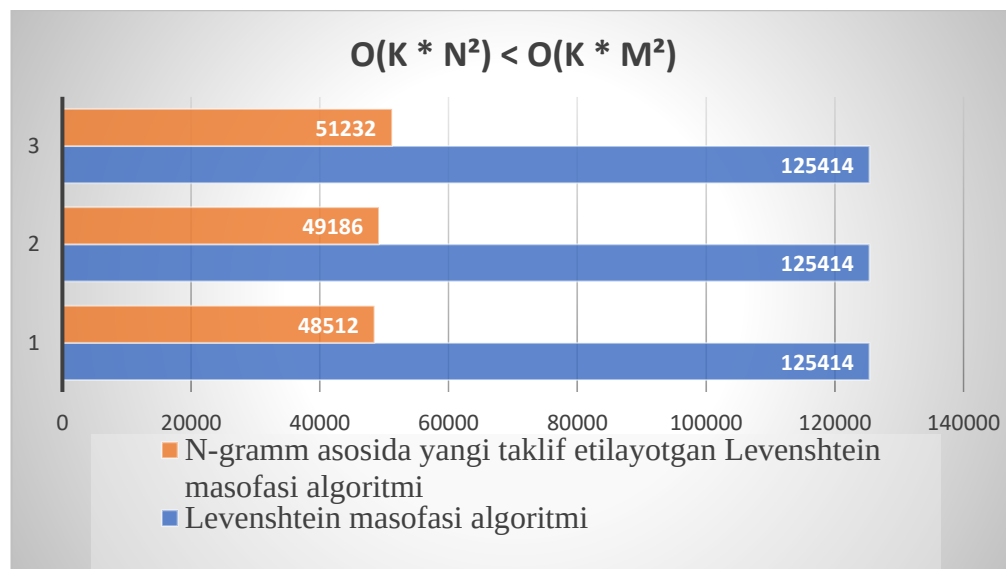
3. Ikkita ro'yxat birlashtirilgandan so'ng, birlashgan ro'yxat bilan qidiruv satrining tartiblangan ro'yxati o'rtasidagi uzunlik farqini tekshiradi va farqi **threshold_max** dan kam bo'lgan ro'yxatlar Levenshtein masofasi algoritmasiga yuboriladi. Tadqiqotda **threshold_max** qiymati to'qqizga teng, chunki bitta harf xatosi uchta trigramga ta'sir qiladi. Agar uchta harf xatosiga ruxsat bersak, eng ko'pi bilan to'qqizta trigram ta'sir qiladi.

Amaliy natijalar

Amaliy tadqiqot natijalari 1-rasmda keltirilgan. Natijalarda jiddiy farqlar kuzatildi. Tadqiqotda biz taklif qilingan algoritm tomonidan sarflangan vaqtni va faqat Levenshtein masofasi algoritmidan foydalanish orqali sarflangan vaqtni o'lchadik.

1-rasmda ko'rsatilganidek, taklif qilingan algoritm Levenshtein masofasi algoritmidan faqat foydalanish natijalariga nisbatan bajarilgan vaqtdan 3 baravar ko'p kamaytirdi.

1-rasmda uchta so'z uchun taqqoslash keltirilgan va har bir taqqoslashda necha marta tahrir masofasi hisoblanganligi va har bir uchun vaqt ko'rsatilgan.



1-rasm. Amaliy tadqiqot natijalari

So'nggi sinov tahlilida Levenshtein masofasi algoritmi 125 414 martada bajarilgan bo'lsa, yangi taklif qilingan algoritmida Levenshtein masofasi algoritmi faqat 48 512, 49 186 va 51 232 martada bajarilgan.

Xulosa

Avtomatik to'g'rilash bugungi kunda ko'plab veb-ilovalarda muhim xususiyatlardan biri bo'lib, foydalanuvchilarning qidiruvlari va yozma matnlarini avtomatik tarzda to'g'rilash imkonini beradi. Ushbu xususiyatni amalga oshirish uchun mavjud Levenshteyn masofa algoritmidan foydalanish mumkin, bu algoritmning funksionali katta bo'lgan ma'lumotlar

bazalarida ishlash vaqtini sezilarli darajada oshiradi. Levenshteyn masofasi algoritmining ishlash davomiyligi $O(M \cdot N^2)$ ni tashkil qiladi, bu yerda M ma'lumotlar bazasidagi so'zlarning sonini, N esa so'zlarning o'rtacha uzunligini bildiradi. Shu sababli, bu algoritmnining ishlashi ko'plab resurslarni talab qiladi.

Shunday ekan, taklif qilingan n -gram indeksidan foydalanuvchi algoritim imlo tuzatuvchi veb-ilovalar uchun juda samarali hisoblanadi. Bu algoritim, avvalambor, ma'lumotlar bazasidagi so'zlar bilan qidiruv so'zi o'rtasidagi o'zaro bog'lanishlarni optimal darajada aniqlashga yordam beradi. Yangi so'zni qidirish uchun trigramlarni asos qilib olib, so'zlar o'rtasida moslikni hisoblash jarayonini osonlashtiradi.

Natijada, taklif qilingan algoritim veb-ilovalarda avtomatik to'g'rilash xususiyatini samarali amalga oshirish uchun yechim bo'la oladi.

Foydalanilgan adabiyotlar

1. Jong Yong Kim and John Shawe-Taylor. "Fast string matching using an n -gram algorithm." Software-Practice and Experience 24(1):79-88. January 1994.
2. Klaus U. Schulz and Stoyan Mihov "Fast string correction with Levenshtein automata" INTERNATIONAL JOURNAL ON DOCUMENT ANALYSIS AND RECOGNITION, Volume 5, Number 1, 67-85, DOI: 10.1007/s10032-002-0082- 8,2003.
3. V. Hodge and J. Austin, "A Comparison of a Novel Spell Checker and Standard Spell Checking Algorithms Pattern Recognition", vol. 15 no. 5, pp. 1073-1081, 2003.
4. Zhao, Zuo-Peng, Yin, Zhi-Min, Wang, Qian-Pin, Xu, Xin-Zheng, Jiang, Hai-Feng and Jisuanji Yingyong "An improved algorithm of Levenshtein Distance and its application in data processing" Journal of Computer Applications. Vol. 29, no. 2, pp. 424-426. Feb. 2009 indepdant.
5. Botir Elov, Alayev Ruhillo, Abdulla Abdullayev, Narzillo Aloyev. "Yuqori N -gram modellarini o'zbek tili matnlariga qo'llash" Raqamli Transformatsiya va Sun'iy Intellect ilmiy jurnali. vol. 2, pp.152-162, OCTOBER 2024.

TEZAURUS VA ONTOLOGIYADA KONSEPT FARQLANISHI TIBBIY TERMINLAR MISOLIDA

Qodirova Zebo

O'zbekiston Milliy universiteti
Kompyuter lingvistikasi va amaliy tilshunoslik
kafedrası o'qituvchisi

Annotatsiya: Hozirgi kunda kompyuter lingvistikasi, axborot texnologiyalari sohasida ko'plab yangi tushunchalar paydo bo'lib, ular orasida *tezaurus* va *ontologiya* kabi tushunchalar ko'p holatlarda qorishiq ifodalanmoqda. Tezaurus va ontologiyalar ma'lumotlarni saqlash, qayta ishlash va qidirishda, shuningdek, sun'iy intellekt va tabiiy tilni qayta ishlash sohalarida keng qo'llanilmoqda. Ushbu maqolada ontologik tipdagi tuzilmalar tahlil qilingan, xususan, tezaurus va ontologiyada konsept farqlanishi, konsept namunalari muhokama etilgan.

Kalit so'zlar: ontologiyada konsept, tezaurusda konsept, subsumption, part hood, instantiation, semantik munosabatlar

Аннотация: В настоящее время в области компьютерной лингвистики и информационных технологий появилось много новых концепций, среди которых такие понятия, как тезаурус и онтология, часто используются как взаимозаменяемые. Тезаурусы и онтологии широко используются при хранении, обработке и поиске данных, а также в областях искусственного интеллекта и обработки естественного языка. В статье анализируются структуры онтологического типа, в частности, обсуждаются дифференциация понятий и закономерности понятий в тезаурусе и онтологии.

Ключевые слова: концепция в онтологии, концепция в тезаурусе, включение, часть, инстанциация, семантические отношения

Annotation: Currently, many new concepts have emerged in the field of computational linguistics and information technology, among which concepts such as thesaurus and ontology are often used interchangeably. Thesaurus and ontologies are widely used in data storage, processing and retrieval, as well as in the fields of artificial intelligence and natural language processing. This article analyzes ontological-type structures, in particular, the distinction of concepts in thesaurus and ontology, and discusses concept examples.

Keywords: concept in ontology, concept in thesaurus, subsumption, part hood, instantiation, semantic relations

Olimlar ontologiya tushunchasiga nisbatan berilgan eng maqbul ta'rif T.Gruberga tegishligini ta'kidlashadi. T. R. Gruber ta'rifida tushunchalarning qay tarzda ochiq ifodalanishi ya'ni qanday tushunchalar borligini aniq ko'rsatib berishni nazarda tutadi. Keyinchalik esa V. N. Borst, J. M. Akermanslar T. R. Gruber ta'rifini rivojlantirib "Umumiy tushunchalarning rasmiy ifodalanishi" tarzida ifoda etadilar va rasmiylik va umumiylikka urg'u beriladi. Ontolog olim M. Uschold tadqiqotlarida ontologiyada bo'lishi kerak bo'lgan asosiy tarkibiy qismlarni ajratib ko'rsatadi:

atamalar lug'ati (termin),
terminlarning ma'nolarini aniq belgilash,
tushunchalar o'rtasidagi bog'lanishlardan iborat ekanligi qayd etiladi.

T. Berners-Lee, V.Hall, J.Hendler, N.Shadbolt, D.Veitznerlar esa ontologiya rasmiy va norasmiy shaklda bo'lishi mumkinligini qayd etishgan. Formal ontologiyaning afzalligi sifatida mashinalar o'qiy oladigan formatda bo'lishi va bu esa semantik veb uchun samarali ishlov berish imkonini yaratishi ta'kidlanadi. P.Green M.Rosemannlar ontologiya turli tushunchani ifodalashini ta'kidlab, ko'lam, darajasi, va maqsadiga ko'ra turlarini farqlashgan.

L.Zemmouchi-Ghomari, A.Ghomarilar tadqiqotida esa ontologiya va terminologiyadagi farqlarni tahlil qilingan. Mazkur tadqiqotda ontologiya konseptlar va ularning o'zaro aloqalari haqida ma'lumot beradigan struktura sifatida ta'riflanib, konseptlar ikki qismga ajratiladi.

1. Superkonsept – bu ma'lum bir konseptni belgilovchi yuqori darajadagi konseptdir.

2. Subkonseptlar esa yuqori darajadagi konseptga bog'liq bo'lgan, lekin o'zining maxsus xossalriga ega bo'lgan konseptlar sifatida ta'riflanadi [L.Zemmouchi-Ghomari, A.Ghomarilar, 2013.315-331]. Qolaversa, konseptlar o'rtasida bir qator munosabat turlari ham o'z ifodasini topadi. L.Zemmouchi-Ghomari, A.Ghomarilar tadqiqotida konseptlar o'rtasidagi asosiy munosabatlarni sanab o'tadilar, quyida ulardan ayrimlarini keltiramiz.

Subsumption (iyerarxik tuzilma) – bu ontologiya va terminologiyada asosiy munosabat turlaridan biridir. Bu munosabatda bir konsept (subkonsept) boshqa bir konsept (superkonsept) ichiga kiritiladi. Boshqacha qilib aytganda, *subsumption* munosabati orqali bir konsept boshqa bir konseptning maxsus birligi sifatida ifodalanadi. *Subsumption* konseptlar o'rtasidagi iyerarxik munosabatni ko'rsatadi. Masalan, “Hayvon” superkonseptiga “mushuk” va “it” subkonseptlari kiradi. Qolaversa, *subsumption* orqali bir konseptning xususiyatlari, superkonseptning xususiyatlariga bog'liq ekanligini ifodalaydi ya'ni “Hayvon” konsepti “Nafas olish” xususiyatiga ega bo'lsa, “mushuk” va “it” ham ushbu xususiyatga ega ekanligi ma'lum bo'ladi.

Part hood (qism munosabati) –bu ontologiya va terminologiya doirasida bir konseptning boshqa bir konseptning qismi sifatida ifodalaydi ya'ni qism va butun konseptlar o'rtasidagi aloqani ko'rsatadi. *Parthood* munosabatida bir konsept (qism) boshqa bir konseptning (butun) tarkibiy qismi sifatida beriladi. Masalan, “Yer” konsepti “okean” va “quruqlik” kabi qism konseptlariga ega. Agar konsept qism bo'lsa, u holda butun konseptning xususiyatlarini ham aks ettirishi mumkin. *Parthood* munosabatlari ko'pincha dinamik bo'lishi ya'ni vaqt o'tishi bilan o'zgarishi yoki yangi qism konseptlar qo'shilishi mumkin.

Connection (bog'lanish munosabati) – bu ontologiyada konseptlar o'rtasida o'zaro aloqani ifodalaydi. Bu munosabat, bir konseptning boshqa bir konsept bilan qanday bog'liq ekanligini ifodalaydi.

Instantiation – bu ontologiya va terminologiyada bir konseptning (sinf yoki kategoriya) aniq bir obyektga (indivud) mos kelishini ifodalaydi. Bu munosabat, nazariy konseptning amaliy misolda qanday ifodalanishini ko'rsatadi. *Instantiation* munosabatida bir obyekt (masalan, “Mening mushugim”) ma'lum bir sinfga (masalan, “Mushuklar”) tegishligini ifodalaydi.

Yuqoridagi tahlil natijasida quyidagi ta'rifni keltirish mumkin: ontologiya – bu obyektlar, hodisalar, tushunchalar va ularning o'zaro bog'lanishlarini aniqlab beradigan va tizimlashtiradigan modeldir. Ontologiyalar semantik munosabatlarni ifodalash bilan birgalikda, munosabatlarni aniq aksiomalar asosida belgilaydi. Ontologiya ma'lum bir sohadagi barcha tushunchalar va ularning o'zaro munosabatlarini ifoda etadi. Ontologiyaning asosiy xususiyati shundaki, u nafaqat so'zlar o'rtasidagi aloqalarni belgilaydi, balki ushbu aloqalar va tushunchalar o'rtasidagi murakkab bog'lanishlarni aniqroq tushunishga imkon yaratadi. Ontologiyada klassifikatsiya, iyerarxiya, atributlar va xususiyatlar hamda o'zgaruvchilar haqida aniqroq tavsiflar mavjud bo'ladi.

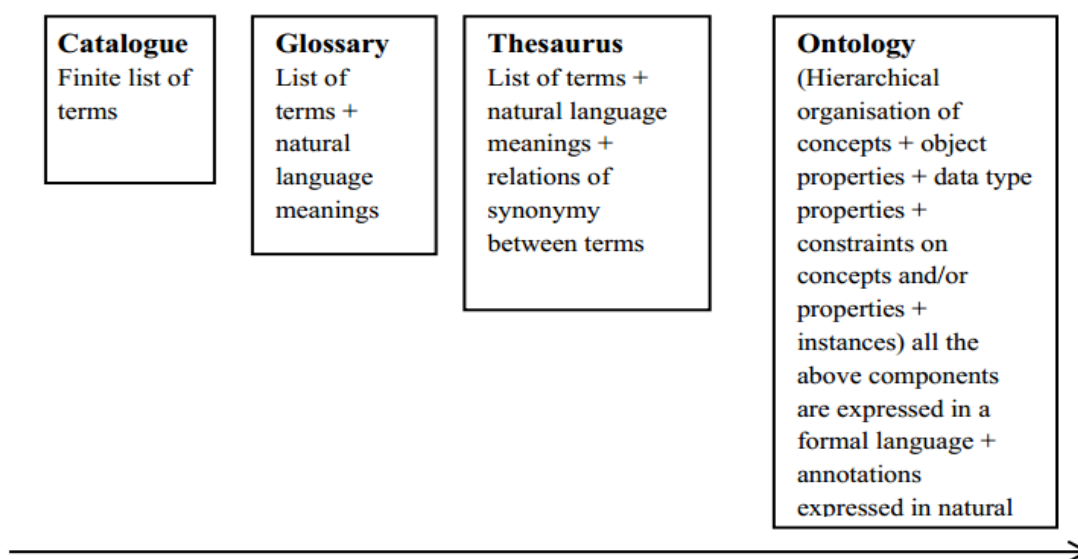
Konsept termini tilshunoslikda o'tgan asrning 80-yillariga qadar tushuncha so'ziga sinonim sifatida ishlatilgan bo'lsa, bugungi kunga kelib uning izohi tushuncha terminiga nisbatan kengayganini ko'rish mumkin [Dimyankov V.Z., 2001. 35-47]. Anglashiladiki, konsept turli semantik texnologiyalarda qanday birliklar bilan ifodalanishida farqlar yuzaga keladi. Quyida farqlarni ko'rib chiqamiz.

Leksik birliklar o'rtasidagi semantik munosabatlarni aniqlashga qaratilgan texnologiyalardan biri *tezaus* sifatida qaraladi. Tezaus va ontologiya o'rtasidagi farqni aniqlash uchun dastlab tezaus va ontologiyaning tarkibiy tuzilishi hamda asosiy munosabatlari tahlilini amalga oshirish lozim.

L.Zemmouchi-Ghomari, A.Ghomarilar (1-rasmda) ontologik tuzilmalarni to'rtta bosqichga bo'lishgan:

1. Catalogue (katalog) – so'zlar ro'yxati, faqat terminlar ro'yxati mavjud.

2. Glossary (izohli lug‘at) – terminlar ro‘yxati va ularning tabiiy tildagi ma‘nolari berilgan.
3. Thesaurus (tesaurus) – terminlar ro‘yxati + tabiiy til ma‘nolari + sinonimlik munosabatlari ya‘ni so‘zlar orasidagi o‘zaro bog‘liqliklar ham ko‘rsatilgan.
4. Ontology (ontologiya) – murakkab tuzilma:
konseptlar (tushunchalar) iyerarxiyasi
obyekt xususiyatlari
ma‘lumotlar turi xususiyatlari
konseptlar va xususiyatlarga cheklovlar
namunalari (instances)
formal tilda ifodalanadi
qo‘shimcha ravishda tabiiy tilda annotatsiyalar beriladi



1-rasm. L.Zemmouchi-Ghomari, A.Ghomarilarning Ontologiyaning rasmiylashtirilishi va murakkabligiga ko‘ra ontologiya spektri

1-rasmda, terminologiya va ontologiya yaratish jarayonlari solishtirilgan. Ontologiya tuzilishi jihatdan murakkab bo‘lib, ontologik tipdagi tuzilmalarning murakkablik darajalari oddiy ro‘yxatdan boshlab to to‘liq formal ontologiyagacha bo‘lgan spektri ifodalangan.

Tezaurus – bu ma‘lum bir soha yoki tilga oid so‘zlar o‘rtasidagi sinonimlik, antonimlik kabi semantik bog‘lanishlarni ifodalaydi. Tezaurusda asosiy e‘tibor ma‘lum bir so‘zlar o‘rtasidagi aloqalarni o‘rnatishga qaratilgan bo‘lib, ularni bir-biriga bog‘lash orqali ma‘lumotlarni ko‘proq strukturaviy formatda taqdim etadi. Tezaurus tushunchalariga xos munosabat turlari: sinonimiya, iyerarxik va assotsiatsiyaviylikdir [Solovev V.D., Dobrov B.V., Ivanov V.V., Lukashevich N.V., 2006. 157]. Tezauruslarda asosiy birlik atamalar bo‘lib, ular deskriptorlar (tavsiflovchilar), askriptorlar (tavsiflovchi bo‘lmaganlar)ga bo‘linadi [Lukashevich N.V., Dobrov B.V., 2015. 47-69]. O‘z mohiyatiga ko‘ra, deskriptorlar mavzu sohasi tushunchalariga mos keladi. Ma‘lum sohada tushunchalar bir nechta muqobil variantga ega deb qaralishi mumkin va variantlar orasidan deskriptorlar tanlanadi, bu esa tezaurus tarkibidagi tushunchaga murojaat qilish usuli sifatida qayd etiladi. Askriptorlar esa deskriptorlarni topishda yordamchi vositalar yoki deskriptor bo‘lmaganlar sifatida qaraladi.

Konsept tushunchasi har ikkala tizimda ham mavjud, lekin ularning ta‘rifi va ishlatilishi farq qiladi. Tezaurusda konsept, asosan, so‘zlar o‘rtasidagi sinonimik yoki antonimik aloqalar sifatida tushuniladi. Bu yerda konsept ko‘proq tilga oid bo‘lib, so‘zlar o‘rtasidagi bog‘lanishlar orqali ifodalanadi. Ontologiyada esa konsept ko‘proq ma‘lumotlarni va ularning o‘zaro aloqalarini chuqurroq tushunishga imkon beradigan tizim sifatida ta‘riflanadi. Ontologiyada konsept

faqatgina soʻz yoki ibora emas, balki bu soʻzning taʼrifi, xususiyatlari, atributlari va boshqa tushunchalar bilan bogʻliqligida namoyon boʻladi.

Ontologiyada konseptlarning farqi, shuningdek, ularning iyerarxik tizimida ham aks etadi. Tezaurusda konseptlarning iyerarxik bogʻlanishlari odatda soddaroq va maʼlum bir soʻzning sinonimlari bilan chegaralanadi. Ontologiyada esa konseptlar bir-biri bilan murakkab iyerarxik tizimda bogʻlanadi.

Tibbiy terminlar tezaurusi namunasi: Tezaurusda asosiy eʼtibor terminlar oʻrtasidagi sinonimlik, antonimlik munosabatiga qaratiladi.

Tibbiyot (sohasi)

Termin naʼmunasi: kasallik

Sinonimi: betob, xasta, illat, dardli, nosogʻ, shikasta, nosogʻlom

Antonimi: baquvvat, farovonlik, kuchli, tuzalmoq, tiklanmoq, dadillashmoq, sogʻayib ketish, tetiklik

Assotsiyalashgan: zaiflik, ogʻriq, isitma, nimjonlik, harorat, kuchsizlanish, darmonsizlanish, madorsizlik, quvvatsizlik, shikoyat

B.Suyunov tadqiqotida tezauruslar zamonaviy kompyuter lugʻati boʻlib, anʼanaviy lugʻatlarga qaraganda keng imkoniyatlarga ega deya tavsiflanadi [Suyunov B.T., 2022: 241]. Shuningdek, olim tadqiqotida “Tezaurusning bugungi terminologiyada zarurat ekanligini quyidagicha asoslash mumkin: tezarus sinonimiya bilan bogʻliq muammoni bartaraf etadi, tushunchaga yaqin boʻlgan soʻzlarni qidirish imkonini kengaytiradi, maʼlumotlarning aniqligini taʼminlaydi” deya izohlanadi. Haqiqatdan ham tezauruslar bugungi kunda axborot qidirish sohalarida sunʼiy intellekt tomonidan faol foydalanish imkoniyatini beradi.

Tibbiyot ontologiyasi namunasi: Ontologiyada tibbiy tushunchalar va kasalliklar oʻrtasidagi bogʻlanishlar quyidagicha ifodalanadi.

Tibbiyot (umumiy kategoriya)

Kardiologiya (subkategoriya)

Pediatriciya (subkategoriya)

Dietologiya (subkategoriya)

Oftalmologiya (subkategoriya)

Stomatologiya (subkategoriya)....

Kardiologiya (subkategoriya)

Kardiologik kasalliklar (subkategoriya)

Ishimik yurak kasalliklari

Stenokardiya

Miokard infarkti

Yurak yetishmovchiligi

Yurak ritmi va oʻzgaruvchanligining buzilishi

Taxikardiya

Ekstrasistoliyalar

Gipertoniya va qon bosimi bilan bogʻliq kasalliklar

Arterial gipertoniya

Gipotoniya....

Tezaurus va ontologiya oʻrtasidagi farqlar konsept tushunchasining ishlatilishi va tasniflanishida aks etadi. Tezaurusda konseptlar asosan soʻzlar oʻrtasidagi semantik aloqalar orqali tushunilsa, ontologiyada konseptlar yanada kengroq va chuqurroq tizimda, iyerarxik aloqalar va atributlar bilan birga aniqlanadi.

Xulosa qilish mumkinki, tezaurus va ontologiyalar maʼlumotlarni tuzishda va ulardan foydalanishda oʻziga xos ahamiyat kasb etadi. Tezaurusda konseptlar oʻrtasidagi bogʻlanishlar soʻzlar va ularning sinonimlari yoki antonimlari orqali amalga oshiriladi. Bu bogʻlanishlar soddaholatda ifodalanadi va maʼlum bir soʻzning oʻxshash yoki qarama-qarshi maʼnolari koʻrsatiladi. Ontologiyada esa konseptlar kengroq iyerarxik tuzilmalarga ega boʻlib, ular bir-biri bilan

murakkab aloqalar orqali bog‘lanadi. Har bir konseptning o‘ziga xos xususiyatlari, attributlari va o‘zaro munosabatlari aniq belgilanadi.

Foydalanilgan adabiyotlar:

1. Mirsaid, A., Bibigul, R., Altinbay, S., & Nilufar, A. (2018). Ontology of grammar rules as example of Noun the Uzbek language and Kazakh languages VI international scientific conference" Modern problems of the applied mathematics and information technology. Al Khorezmiy-2018.
2. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
3. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
4. Suyunov B. T. Tibbiy terminlar semantikasi va tezaurusi: diss. ... Filol. fanlari doktori – (Dsc) Toshkent, 2022. – 241 b.
5. Демьянков В.З. Понятие и концепт в художественном литаратуре и в научном языке // Вопросы филологии. –Москва, 2001. -№ 1. –С. 35-47.
6. Лукашевич Н.В., Добров Б.В. Проектирование лингвистических онтологий для информационных систем в широких предметных областях. (2015). 1(15), 47–69.
7. Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В. Онтологии и Тезаурусы. Москва. 2006. 157 с.
8. Zemmouchi-Ghomari L., Ghomari A., “Terminologies versus Ontologies fromtheperspective of ontologists,” Int. J. Web Sci., 2013. – P 315-331.

O'ZBEK TILI GRAMATIKASIDAGI SO'ZLARNI TURKUMLARIGA AJRATISHDA NLP MODELLARIDAN FOYDALANILANISH

Nazirova E.Sh.

DsC. Professor

Muhammad al-Xorazmiy nomidagi
Toshkent Axborot Texnologiyalari Universiteti

Abidova Sh.B.

PhD. v.b.dotsent

Muhammad al-Xorazmiy nomidagi
Toshkent Axborot Texnologiyalari Universiteti

Xolmuminov Y.X.

PhD doktorant

Raqamli texnologiyalar va sun'iy
intellektni rivojlantirish ilmiy-tadqiqot institute

Anotatsiya. O'zbek tili grammatikasidagi so'zlarni turkumlarga ajratish (POS-tagging) tabiiy tilni qayta ishlashning (NLP) asosiy vazifalaridan biridir. Bu jarayon, tilning sintaktik va morfologik strukturasi haqida chuqur ma'lumot olish uchun zarur bo'lib, so'zlarni to'g'ri tahlil qilishni ta'minlaydi. O'zbek tilida so'z turkumlarini aniqlash, uning morfologik va sintaktik xususiyatlarini hisobga olgan holda, maxsus yondashuvlar va modellarni talab qiladi. Ushbu maqolada, o'zbek tilida so'zlarni turkumlarga ajratish uchun qo'llaniladigan zamonaviy NLP modellari tahlil qilinadi. Maqola doirasida an'anaviy qoidalarga asoslangan yondashuvlar, statistik modellar va chuqur o'rganish (deep learning) usullaridan foydalanib, o'zbek tilida so'z turkumlarini aniqlash jarayoni ko'rib chiqiladi. Bunda, HMM (Hidden Markov Model), CRF (Conditional Random Fields) va boshqa transformator asosidagi modellar o'zbek tiliga qanday moslashtirilgani tahlil qilinadi. Shuningdek, morfologik murakkabliklar, leksik resurslarning roli, va o'zbek tilining erkin so'z tartibi kabi omillar, POS-taggingning samaradorligini ta'sir qiluvchi faktorlar sifatida muhokama qilinadi. Maqola, o'zbek tilida so'z turkumlarini aniqlashda mavjud bo'lgan texnik muammolarni yoritish va ularni hal qilish uchun zamonaviy NLP yondashuvlarining imkoniyatlarini ko'rsatishni maqsad qiladi. Shuningdek, tilni qayta ishlash texnologiyalarini yanada rivojlantirish uchun ilmiy hamjamiyatga yangi yondashuvlar taklif etiladi.

Kalit So'zlar. HMM (Hidden Markov Model), CRF (Conditional Random Fields), POS-tagging, morfologik, NLP, deep learning.

Аннотация. Определение частей речи (POS-tagging) в грамматике узбекского языка является одной из ключевых задач обработки естественного языка (NLP). Этот процесс необходим для глубокого анализа синтаксической и морфологической структуры языка и обеспечивает правильный анализ слов. В статье рассматриваются современные модели NLP, используемые для классификации слов в узбекском языке. Основное внимание уделено моделям HMM, CRF и трансформерным архитектурам, их адаптации под узбекский язык, а также сложности морфологии, роли лексических ресурсов и свободного порядка слов. Цель статьи — осветить существующие проблемы и предложить новые подходы для совершенствования NLP-инструментов на узбекском языке.

Ключевые слова: HMM, CRF, POS-tagging, морфология, NLP, глубокое обучение.

Abstract. Part-of-speech tagging (POS-tagging) in Uzbek language grammar is a fundamental task in Natural Language Processing (NLP). This process is essential for extracting

syntactic and morphological structure and enabling accurate word analysis. This article analyzes modern NLP models applied to Uzbek, including rule-based, statistical, and deep learning approaches. It focuses on the adaptation of HMM (Hidden Markov Model), CRF (Conditional Random Fields), and transformer-based models to Uzbek linguistic features. Furthermore, morphological complexity, the role of lexical resources, and the free word order of Uzbek are discussed as influential factors. The paper aims to highlight technical challenges in POS-tagging for Uzbek and explore advanced NLP solutions to address them.

Keywords: HMM, CRF, POS-tagging, morphology, NLP, deep learning.

Kirish. O'zbek tili, o'zining boy grammatikasi va morfologik tuzilishi bilan ajralib turadi. Tilshunoslikda so'zlarni to'g'ri turkumlarga ajratish (POS-tagging) tahlili morfologik tahlilning asosiy bosqichlaridan biridir. So'z turkumlarini aniqlash, matnni qayta ishlash jarayonida, ayniqsa tabiiy tilni tushunish va ma'lumotlarni tahlil qilishda muhim ahamiyatga ega. Tabiiy tilni qayta ishlash (NLP) texnologiyalari yordamida so'z turkumlarini aniqlashning samarali va aniq usullarini ishlab chiqish mumkin. O'zbek tili uchun so'z turkumlarini aniqlashda, bir nechta ilg'or yondashuvlar va modellar mavjud. HMM (Hidden Markov Model), va CRF (Conditional Random Fields) kabi ilg'or sun'iy intellekt modellaridan foydalanish bu jarayonni yanada samarali qilishga yordam beradi. HMM va CRF kabi statistik modellar, xususan, so'zlarni va ularning turkumlarini o'zaro bog'liq holda tahlil qilishga yordam beradi. Mazkur maqolada, O'zbek tili grammatikasidagi so'zlarni turkumlariga ajratish jarayonida qo'llaniladigan HMM va CRF modellarining samaradorligi va ularning bir-biriga nisbatan afzalliklari tahlil qilinadi. Ushbu modellar yordamida so'z turkumlarini aniqlash jarayonini yanada samarali va aniq qilish mumkin, bu esa o'z navbatida turli NLP vazifalarini bajarishda, masalan, tarjima, matnni tahlil qilish va intellektual yordam tizimlarida foydalanishga yordam beradi.

Ushbu tadqiqotda o'zbek tilidagi so'zlarning morfologik va sintaktik xususiyatlarini ifodalovchi **dataset.csv** faylidan foydalanildi. Mazkur to'plamda har bir qator quyidagi ustunlardan iborat:

- **So'z** – tahlil qilinadigan real so'z shakli;
- **So'z turi** – tegishli so'z turkumi (masalan: ot, fe'l, sifat);
- **O'zaks** – so'zning grammatik asosini bildiruvchi ildiz shakli;
- **Shakl** – so'zning umumiy morfologik ko'rinishi (yasalgan yoki qo'shimcha olgan shakli);
- **Qo'shimchalar** – qo'llangan grammatik affikslar yoki qo'shimchalar.

Ma'lumotlar 672 ta qatordan iborat bo'lib, bu o'zbek tili morfologiyasini avtomatik tahlil qilish uchun yetarlicha dastlabki asos bo'la oladi.

HMM modeli (Hidden Markov Model). HMM modeli statistik ehtimollar asosida ketma-ketlikdagi yashirin holatlarni aniqlashga qaratilgan bo'lib, tabiiy tilni qayta ishlashda ayniqsa POS-tagging jarayonida keng qo'llaniladi. Model quyidagi tamoyilga asoslanadi:

- Har bir so'z kuzatiladigan hodisa bo'lib xizmat qiladi;
- So'z turkumi esa yashirin holat deb qaraladi;
- Model har bir so'zning ma'lum bir turkumga tegishli bo'lish ehtimolini aniqlaydi.

Ushbu tadqiqotda hmmlearn kutubxonasidan foydalanildi. So'zlar va ularning teglarining sonli ifodalari LabelEncoder yordamida raqamlashtirildi va model MultinomialHMM yordamida o'qitildi.

```
model = hmm.MultinomialHMM(n_components=len(np.unique(y)), n_iter=1000, random_state=42)
model.fit(X_train)
logprob, y_pred = model.decode(X_test, algorithm="viterbi")
```

1-rasm. HMM modeli training qilinish jarayoni.

CRF modeli (Conditional Random Fields). CRF modeli — bu shartli ehtimollik modeli bo‘lib, so‘zlar ketma-ketligidagi har bir token (so‘z) uchun kontekstga asoslangan holatda label (teg) aniqlashga mo‘ljallangan. Bu model ayniqsa o‘zaro bog‘liq bo‘lgan belgilar (features) asosida qaror chiqarishda yuqori aniqlikka ega.

Tadqiqotda sklearn-crfsuite kutubxonasi yordamida CRF modeli qurildi. Har bir so‘z uchun quyidagi xususiyatlar (features) hisobga olindi:

- So‘zning o‘zi (token)
- So‘zning o‘zagi
- Qo‘shimchalari
- Oldingi va keyingi so‘zlar
- Harflar uzunligi va katta/kichik harf tahlili

Model trening uchun ketma-ketlikka asoslangan fit() funksiyasi yordamida o‘qitildi.

```
crf = sklearn_crfsuite.CRF(
    algorithm='lbfgs',
    max_iterations=100,
    all_possible_transitions=True
)
crf.fit(X_train, y_train)
```

2-rasm. CRF modelning training qilinish jarayoni.

Modelni baholash usullari. HMM va CRF modellarini training va test qiligandan so‘ng har bir modelni samaradorligini baholash lozim buning uchun biz bazi usullardan foydalanamiz. HMM va CRF modellarini baholashda quyidagi mezonlardan foydalanildi:

- **Accuracy (aniqlik)** – umumiy to‘g‘ri bashorat qilingan so‘z turkumlarining foizi;
- **Classification Report** – precision, recall va F1-score qiymatlari asosida to‘liq baho;
- **Confusion Matrix** – model tomonidan bashorat qilingan va haqiqiy so‘z turlari orasidagi farqni grafik ko‘rinishda ko‘rsatish.

Accuracy (aniqlik) — bu model nechta to‘g‘ri bashorat qilganligini umumiy bashoratlardan soniga nisbatan o‘lchaydi.

Matematik formulasi:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Bu yerda:

TP — To‘g‘ri ijobiy natijalar (True Positives)

TN — To‘g‘ri salbiy natijalar (True Negatives)

FP — Noto‘g‘ri ijobiy natijalar (False Positives)

FN — Noto‘g‘ri salbiy natijalar (False Negatives)

Precision (Aniqlik)-Model ijobiy deb belgilaganlarning nechta haqiqatan ham ijobiy bo‘lganligini ko‘rsatadi.

$$\text{Precision} = TP / (TP + FP)$$

Recall (Qamrov) — Model barcha haqiqiy ijobiylardan nechta to‘g‘ri topganini bildiradi.

$$\text{Recall} = TP / (TP + FN)$$

F1-score — Precision va Recall o‘rtasidagi harmonik o‘rtacha. Har ikkalasiga teng e‘tibor beradi.

$$\text{F1-score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

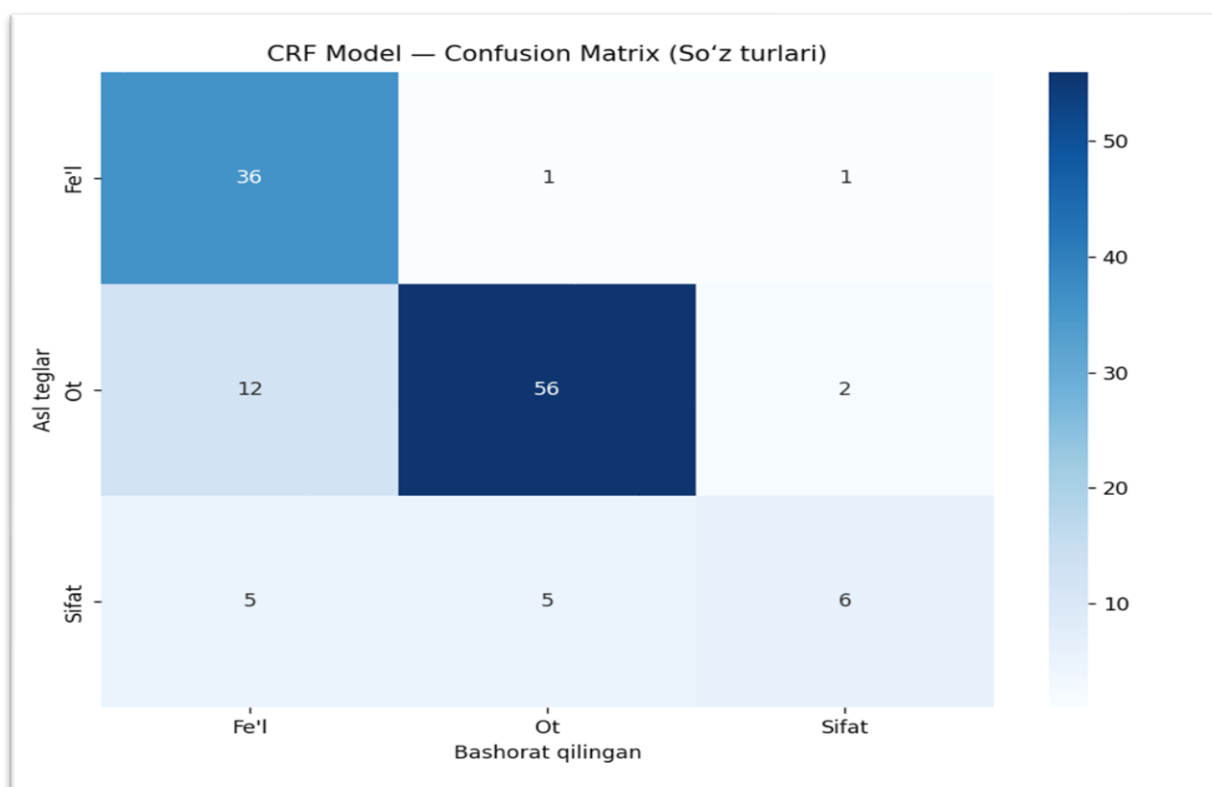
Baholash uchun train_test_split yordamida ma'lumotlar 80% trening, 20% test sifatida ajratildi. Baholash natijalari orqali HMM va CRF modellarining kuchli va zaif tomonlari aniqlab berildi.

```
print(classification_report(y_test, y_pred, target_names=tag_encoder.classes_))
confus_mat = metrics.confusion_matrix(y_test, y_pred)
sns.heatmap(confus_mat, annot=True, fmt='g')
```

3-rasm. Modelni baholash jarayoni

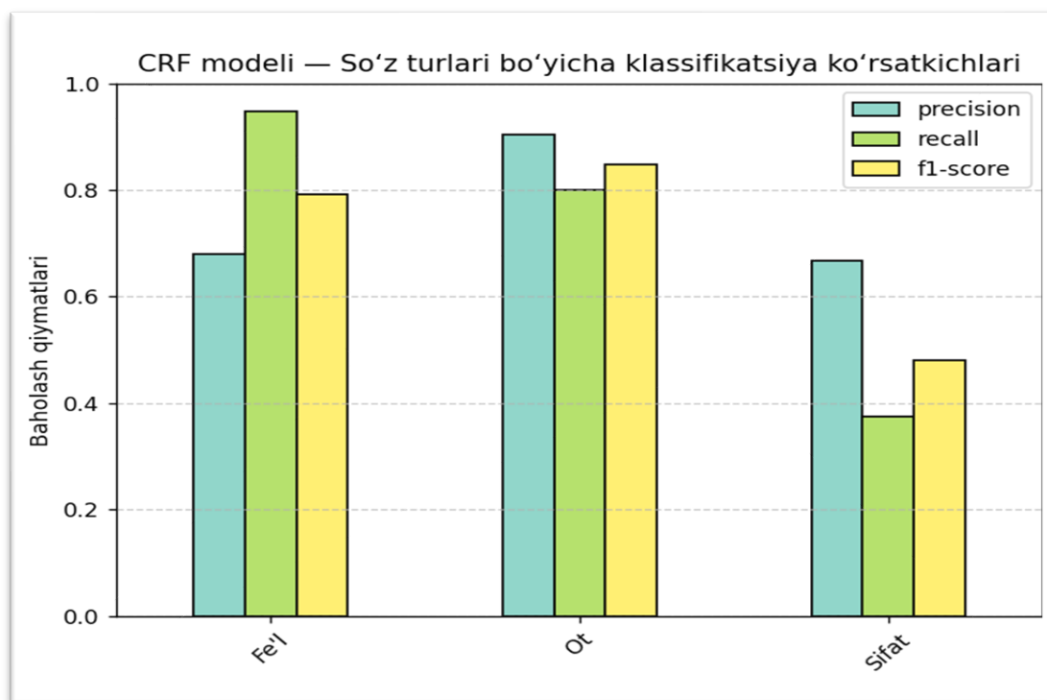
Natijalar. Mazkur tadqiqotda o'zbek tili so'zlarini turkumlariga ajratishda HMM (Hidden Markov Model) va CRF (Conditional Random Fields) modellarining ishlash samaradorligi baholandi. Model natijalari accuracy, precision, recall, va f1-score mezonlari asosida solishtirildi.

CRF modeli asosida olingan natijalari. CRF modeli yordamida 500 ta test namunasi ustida quyidagi natijalar kuzatildi:

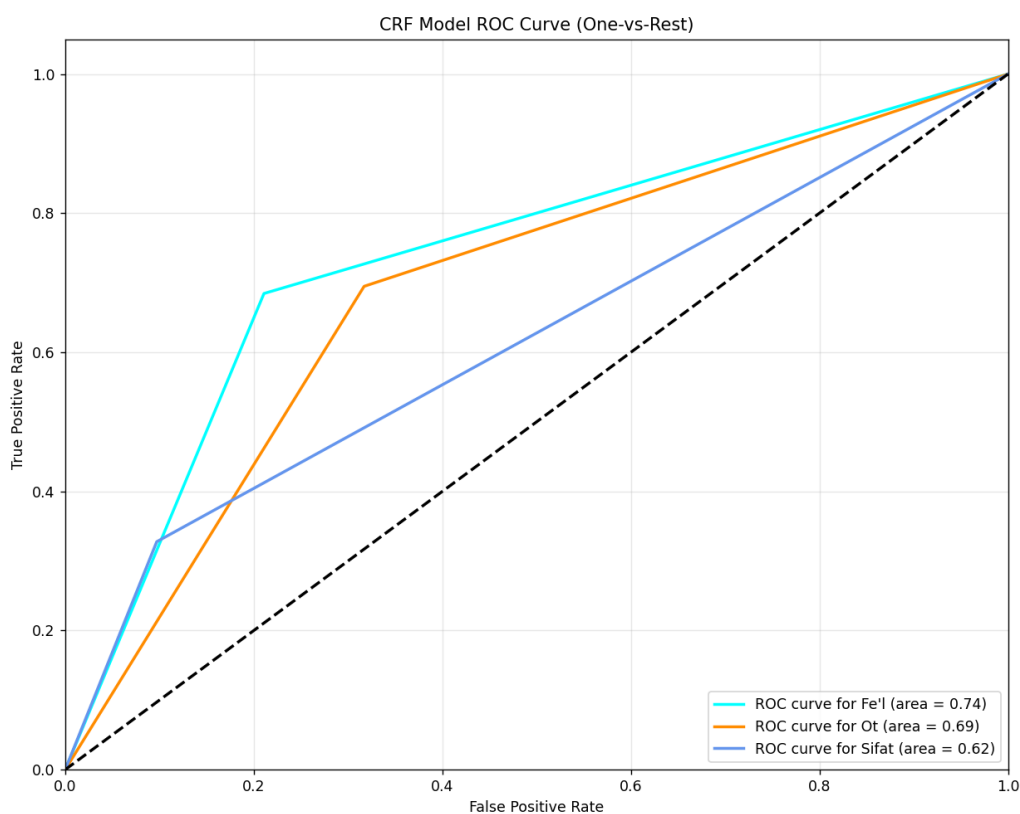


4-rasm. Crf modeling test qilingandagi natijasi.

- Umumiy aniqlik (accuracy): 0.79
- Ot va fe'l so'z turkumlarida modelning precision va recall ko'rsatkichlari o'rtacha darajada chiqdi:
 - Ot: precision = 0.619, recall = 0.695
 - Fe'l: precision = 0.622, recall = 0.685
- Sifat so'z turkumi bo'yicha modelning natijalari ancha past: precision = 0.513, recall = 0.328
- Bunday pastlikka sabab sifatga tegishli misollar sonining kamligi yoki ularning o'zaro o'xshashligi bo'lishi mumkin.

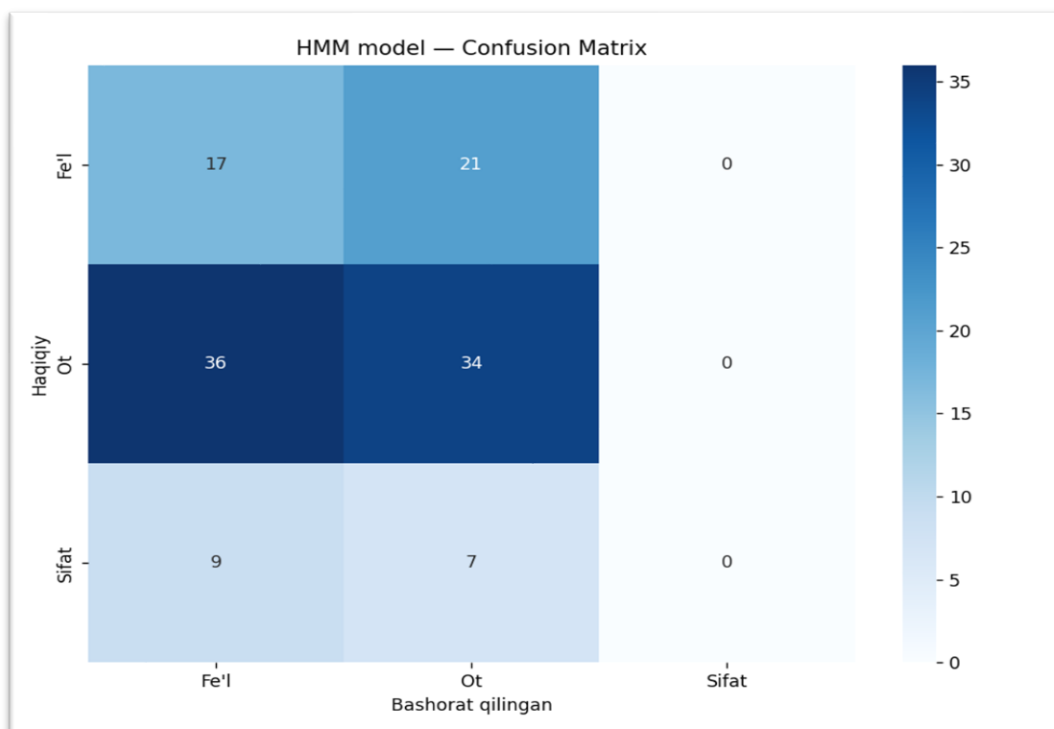


5-rasm. CRf modeling aniqligi grafig koʻrinishda



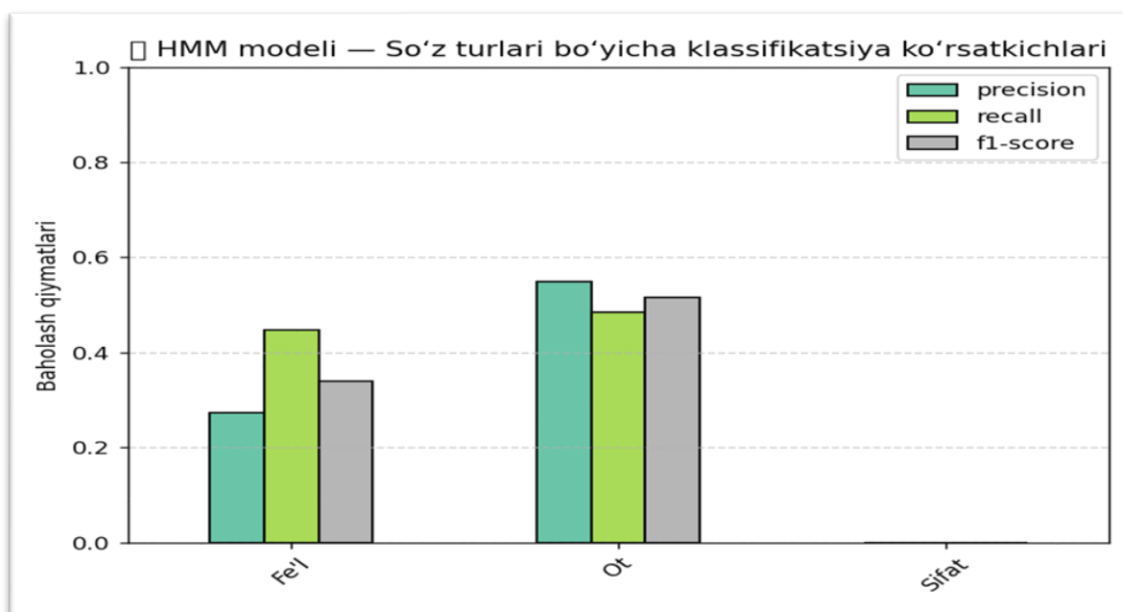
6-rasm. Crf modeling ROC egri chizigʻi

HMM modeli asosida olingan natijalari. HMM modeli esa 20 ta test namunasi ustida quyidagi natijalarni koʻrsatdi:

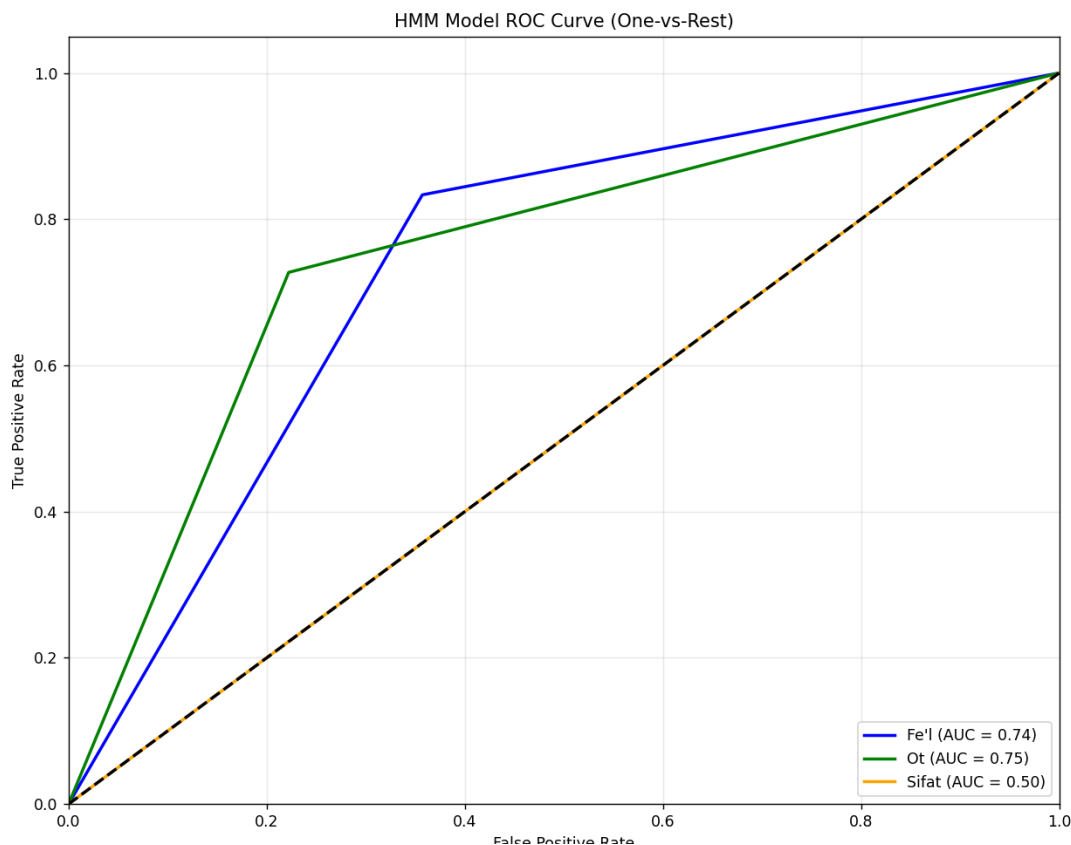


7-rasm. Hmm modelining test qilingandagi natijasi.

- Umumiy aniqlik (accuracy): 0.65
- Ot va fe'l so'zlar uchun nisbatan yaxshi aniqlik qayd etildi:
 - Ot: f1-score = 0.76
 - Fe'l: f1-score = 0.62
- Biroq sifat so'z turkumi bo'yicha umuman aniqlik bo'lmadi (precision = 0.00, recall = 0.00).
- HMM modeli kontekst va strukturaviy elementlarni chuqur o'rganishga mos emas, natijada noaniq yoki kam uchraydigan so'zlar (xususan, sifatlar) noto'g'ri tasniflangan.



8-rasm. Hmm modeling aniqligi grafig ko'rinishda



9-rasm. Hmm modeling ROC egri chizig'i

Xulosa

Tahlil natijalariga ko'ra, CRF va HMM modellarining har biri o'ziga xos afzallik va cheklavlarga ega ekani aniqlandi. CRF modeli kontekst va strukturaviy xususiyatlarni chuqurroq tahlil qilishga qodir bo'lib, ayniqsa ot va fe'l so'z turkumlarini aniqlashda yuqori aniqlikka erishdi. Bu modelning f1-score ko'rsatkichlari amaliy jihatdan qoniqarli bo'lib, uni POS-tagging vazifalari uchun foydali vosita sifatida tavsiya qilish mumkin. HMM modeli esa statistik yondashuvga asoslanganligi sababli o'zbek tilining murakkab morfologik tuzilmalari, ayniqsa sifat so'zlarini tahlil qilishda qiyinchiliklarga duch keldi. Bu esa HMM modelining kontekst va semantik aloqalarni chuqur o'rgana olmasligidan dalolat beradi. Xulosa qilib aytganda, CRF modeli real amaliyotda qo'llash uchun samaraliroq vosita bo'lib, o'zbek tili uchun so'z turkumlarini aniqlashda HMM modeliga qaraganda ancha ishonchli natijalarni taqdim etadi. Kelajakda bu modellar ustida qo'shimcha ishlov berish, ayniqsa ma'lumotlar sonini ko'paytirish va kontekstga asoslangan zamonaviy transformer modellar bilan solishtirish tavsiya etiladi.

Foydalanilgan adabiyotlar.

1. A. Yusupova, "O'zbek tilida tabiiy tilni qayta ishlashning asosiy masalalari", O'zMU Ilmiy axborotlari, 2021.
2. Lafferty, J., McCallum, A., & Pereira, F. (2001). "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data." In Proceedings of the 18th International Conference on Machine Learning (ICML).
3. Rabiner, L. R. (1989). "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." Proceedings of the IEEE, 77(2), 257–286.
4. Sutton, C., & McCallum, A. (2012). "An Introduction to Conditional Random Fields." Foundations and Trends in Machine Learning, 4(4), 267–373.
5. Jurafsky, D., & Martin, J. H. (2021). "Speech and Language Processing" (3rd ed.). Draft chapters. Stanford University.

6. Bird, S., Klein, E., & Loper, E. (2009). "Natural Language Processing with Python." O'Reilly Media.
7. S. Jo'rayev, "Zamonaviy lingvistikada sun'iy intellekt modellarining o'rni", Filologiya fanlari jurnali, 2020.
8. N. Karimova, "O'zbek tilidagi so'z turlarini avtomatik aniqlash masalalari", O'zbekiston Milliy universiteti nashri, 2022.
9. Sh. Qurbonov, "Lingvistik modellashtirish: nazariya va amaliyot", Toshkent, 2019.
10. R. Hasanov, "Morfologik tahlil va so'z turkumlariga ajratish algoritmlari", Ilmiy-texnik axborotlar jurnali, 2021.
11. Jurabekova M., "O'zbek tilining kompyuter lingvistikasi uchun leksik asoslari", 2020.
12. Nuriddinov D., "O'zbekcha NLP modellarining rivojlanish istiqbollari", Texnologiyalar olami, 2022.
13. S. Saidova, "O'zbek tili morfologik analizatorlari: tahlil va istiqbol", Ilmiy maqolalar to'plami, 2023.
14. Hasanboyev T., "NLP asosida o'zbek tili grammatikasini o'rganish metodlari", O'zMU axborotlari, 2022.
15. Muminova S., "Sun'iy intellekt va tilshunoslik integratsiyasi", Filologiya va amaliy tilshunoslik jurnali, 2021.

O'GZAKI NUTQ KOMPYUTER KORPUSLARINING AHAMIYATI VA UNI TASHKIL QILISH JARAYONI

Umirzaqova Maxina Anvarjon qizi

O'zMU Kompyuter lingvistikasi

yo'nalishi 1- bosqich magistranti

E-mail: mahinasabirzanov@gmail.com

Orcid 0009-0001-6355-7645

Xoljurayeva Yulduz Sobir qizi

O'zMU Kompyuter lingvistikasi

yo'nalishi 1- bosqich magistranti

E-mail: abdurahimovayulduz4@gmail.com

Orcid 0009-0008-8120-8239

Annotatsiya: Maqolada og'zaki nutq kompyuter korpuslarining ahamiyati va uning hozirgi zamonaviy tilshunoslik, sun'iy intellekt va tabiiy tilni qayta ishlash(NLP) sohalaridagi ahamiyati haqida so'z yuritiladi. Shuningdek, og'zaki nutq kompyuter korpuslarining tashkil qilish jarayoni va ularni tashkillashtirishning kamchilik va afzallikari haqida fikr yuritiladi.

Kalit so'zlar: nutq korpuslari, tabiiy tilni qayta ishlash(NLP), spontan nutq, transkripsiya.

Abstract: The article examines the importance of conversational computer bodies and their significance in modern linguistics, artificial intelligence, and natural language processing (NLP). The process of organizing oral computer buildings and the disadvantages and advantages of their organization are also considered.

Keywords: corpus of speech, natural language processing(NLP), spontaneous speech, transcription.

Аннотация: В статье рассматривается важность корпусов разговорных компьютеров и их значение в современной лингвистике, искусственном интеллекте и обработке естественного языка(НЛП). Также рассматривается процесс организации корпусов ЭВМ устной речи и недостатки и преимущества их организации.

Ключевые слова: корпус речи, обработка естественного языка(НЛП), спонтанная речь, транскрипция.

Kirish.Bugungi shiddat bilan rivojlanib borayotgan asrda hayotimizni axborot texnologiyalari , zamonaviy texnikalar va kompyuterlarsiz tasavvur qilish qiyin. Ushbu vositalar hatto, hayotimizning bir bo'lagiga aylanishga ulgurdi desam ham mubolag'a bo'lmaydi, albatta. Bugungi kunni bemalol hech ikkilanmasdan "texnologiyalar asri" deb atay olamiz.

Hozirgi kunda shubhasiz, kompyuter qadami yetmagan hech bir jabha qolmadi,u jamiyatimizning barcha sohalariga bosqichma-bosqich kirib kelmoqda. Natijada ko'plab zamonaviy kasblar va yo'nalishlar yuzaga kelmoqda. Zotan, tilshunoslikda kompyuter texnologiyalaridan uzluksiz va faol ravishda foydalanish ortidan kompyuter lingvistikasi yo'nalishi shakllanishiga zamin yaratdi. Professor B.Mengliyev e'tirof etganidek, kompyuter texnologiyalari sohasi tillarning taraqqiyoti va omon qolishi uchun hal qiluvchi omilga aylanib bo'ldi.

Bugungi kunda jadallik bilan rivojlanayotgan sohalardan biri korpus lingvistikasi tushunchasi tilimizda tez-tez takrorlanmoqda. bo'lmoqda. Umuman olganda, korpus lingvistikasi tilni ilmiy tahlil qilishda kompyuter texnologiyalaridan foydalanadigan soha bo'lib, u til o'rganish ,tilshunoslik va tabiiy tilni qayta ishlash(NLP) sohalarida muhim ahamiyatga ega.

Asosiy qism. Korpus lingvistikasi til qoidalarini aniqlash uchun tilning kata hajmdagi yozma va ogʻzaki nutq shaklidagi namunalarini toʻplaydi va ularni tahlil qiladi. Shuningdek, u tilni chuqur va tizimli oʻrganishga yordam beradi, chunki u tabiiy tilda mavjud boʻlgan barcha xususiyatlarni mavjud boʻlgan maʼlumotlarga tayangan holda oʻrganadi. Ogʻzaki nutq kompyuter korpuslari esa korpus lingvistikasining bir boʻlagi boʻlagi boʻlib xizmat qiladi. Bunday nutq korpuslarida kompyuterlar orqali tahlil qilinadigan, toʻplangan ogʻzaki nutq matn(material)laridan iborat maʼlumotlar toʻplami sanaladi. Bu matnlarga tasodifiy yoki rejalashtirib yozib olingan video yozuvlar, audio matnlar kiradi. Ushbu korpuslar, tabiiy tilni qayta ishlash (NLP) va tilshunoslikda ogʻzaki nutqning oʻziga xos xususiyatlarini aniqlash va oʻrganish uchun muhim vosita hisoblanadi. Ogʻzaki nutq koʻpincha tez-tez toʻxtab qolishlar, noaniq jumlar, tinish belgilarining yoʻqligi, jumlalarining qisqargan shakllari va boshqa xususiyatlar bilan ajralib turadi

Ogʻzaki nutqning kompyuter korpuslari turli xil maqsadlar uchun ishlatiladi

Masalan, tilshunoslik va fonologiyada ogʻzaki nutqdagi tovushlar va ularning fonetik xususiyatlarini oʻrganish uchun. Masalan, soʻzlarning talaffuzi, talaffuzdagi farqlar va boshqalar. Shuningdek, nutqni aniqlash va transkripsiya qilishda audio fayllar asosida nutqni matnga aylantirish va uning toʻgʻri transkripsiyasini yaratish. Bunda, korpusda mavjud boʻlgan ogʻzaki nutq matnlaridan foydalaniladi. Yana bir muhim jihatlaridan biri bu nutqni avtomatik tushunish jarayondir. Ushbu jarayonda kompyuterlar ogʻzaki nutqni avtomatik ravishda tushunishi va unga javob berishi uchun oʻqitiladigan maʼlumotlarni yaratish. Bunda, kompyuter ogʻzaki nutqdagi noaniqliklar va xatolarni aniqlashni oʻrganadi.

Nutq korpusining eng yaxshi tomoni shundaki, bu bizning odatiy suhbatlarimiz haqida yorqin dalillarga ega boʻlishdir. Spontan nutq kutilmagan hodisalarga toʻladir. Biz oʻylagan ibora qanday talaffuz qilinishi bilan kimdir uni haqiqiy suhbatda aytayotganini solishtirish juda qiziqarli. (Collen Richey. Speech Corpora. 2020)

Ogʻzaki nutq korpusining yoki ogʻzaki nutqni oʻrganish uchun yaratilgan maʼlumotlar bazasining bir qator afzalliklari mavjud. Quyida ularning bir nechtasini sanab oʻtamiz.

Avvalo, ogʻzaki nutq korpusi tabiiy tilni oʻrganish uchun muhimdir. Ogʻzaki nutq korpusi, asosan, real hayotda yoki suhbatlar jarayonida ishlatiladigan tilni oʻz ichiga oladi. Bu esa, yozma nutqdan farqli oʻlaroq, tabiiy va unikal nutq uslublarini oʻrganishga imkon beradi.

Yana bir afzalliklaridan biri bu nutqning konteksi hisoblanadi. Ogʻzaki nutq koʻpincha muayyan kontekstda va vaziyatda paydo boʻladi, bu esa nutqning maʼnosi va tuzilishini yaxshiroq tushunishga yordam beradi. Suhbatlar va dialoglar koʻpincha koʻp qatlamli maʼlumotlarni oʻz ichiga oladi, bu esa uning tahlilini boyitadi.

Bundan tashqari, fonetik va prozodik xususiyatlarni oʻrganish orqali koʻzlangan maqsadga yetish mumkin. Ogʻzaki nutqda ovoz, ton, urgʻu va pauzalar kabi prozodik elementlar juda muhim. Korpus orqali bu elementlarni tahlil qilish va oʻrganish mumkin.

Shuningdek, dialekt va leksik farqlarni aniqlash mumkin. Bunda ogʻzaki nutqda turli mintaqalarda, yosh guruhlari yoki ijtimoiy sinflarda oʻziga xos til xususiyatlari va leksik farqlar mavjud. Bunday korpuslar bu farqlarni tahlil qilishga yordam beradi.

Bundan tashqari, suhbatlar va interaktiv nutq jarayonida nutqda kommunikatsiyaning ikki tomonlama jarayoni (javoblar, savollar, soʻrovlar) kuzatiladi, bu esa interaktiv nutqning oʻziga xos xususiyatlarini oʻrganish imkonini yaratadi.

Eng foydali tomonlaridan biri, avtomatik nutq tanish tizimlarini rivojlantirish: Ogʻzaki nutq korpuslari, masalan, nutqni matnga aylantirish (STT – Speech-to-Text) texnologiyalari uchun juda foydalidir. Bu tizimlar koʻproq va aniqroq ishlash uchun keng miqyosdagi ogʻzaki nutq korpuslarini oʻrganishlari kerak.

Shu bilan birga, kulturologik va psixolingvistik tadqiqotlar uchun ham bu usul qoʻl keladi. Ogʻzaki nutqda ijtimoiy va madaniy nuqtayi nazardan koʻplab jihatlar namoyon boʻladi, Bu esa lingvistik va psixolingvistik tadqiqotlarda muhim rol oʻynaydi.

Shu bilan bir qatorda, ogʻzaki nutqni tahlil qilishda baʼzi qiyinchiliklarga duch kelish mumkin, masalan, grammatik noaniqliklar, uzilishlar yoki soʻzlarni toʻgʻri transkripsiya qilishda yuzaga keladigan muammolar mavjud.

Umuman olganda, nutq korpusini ikki turga boʻlish mumkin:

1 Nutqni oʻqish

- Kitoblardan parchalar
- Axborot eshittirishlari
- Soʻzlar roʻyxati
- Raqamlar ketma-ketligi

2 Spontan nutq

- Dialoglar va uchrashuvlar – ikki yoki undan ortiq odam oʻrtasidagi erkin suhbatlar
- Hikoyalar – bir kishi tomonidan hikoya qilinadi
- Xarita-topshiriqlar – ikki kishiga har biriga boshqa odam koʻra olmaydigan xarita beriladi.
- Uchrashuv-vazifalar – ikki kishiga individual jadvallar beriladi va taxmin qilinadi uchrashish uchun boʻsh vaqt topish.

“Oz ustasi” simulyatsiyalari – real hayotdagi vaziyatni modellashtirish, masalan, parvozni bron qilish. (Collen Richey. Speech corpora)

Ogʻzaki nutqning eng mashhur korpuslaridan biri sifatida Britaniya milliy korpusining ogʻzaki nutq korpusini misol qilib olishimiz mumkin. Korpus 1991-yildan 1994-yilgacha boʻlgan davrni oʻz ichiga oladi. Korpus ikki qismdan iborat boʻlib, 1-qism 40% demografik namunalardan olingan, 2-qism 60%i esa kontekstga asoslangan. Korpus asosan ingliz tilida boʻlib, turli ijtimoiy va madaniy guruhlar orasidagi suhbatlar haqida maʼlumot beradi. (N.Z. Abdurahmonova. 2023 , 291-b)

Britaniya ingliz tilining yangi nutq korpusi esa Kembrij universitetida qayd etilgan. Wall Street Journal matn korpusi(WSJCAM0) ingliz tilida soʻzlashuvchi eng yirik korporatsiyalardan birini tashkil qiladi. Maʼlumotlar bazasi har biri 110 ga yaqin nutq soʻzlaydigan 140 ta maʼruzachilardan iborat. (WSJCAM0: A British English speech corpus for large vocabulary continuous speech recognition) .

Britaniya milliy korpusi text encoding initiative (TEI)asosida matnlarni raqamli shaklda kodlagan va shu yoʻriqnoma ogʻzaki nutq aktining struktural, kontekstual annotatsiyadan iborat boʻlib, muayyan tadqiqot maqsadi uchun moslashtirilgan. Maʼlumot uchun TEI, asosan, adabiy va ilmiy matnlarni aniq va samarali shaklda saqlash, tahlil qilish va almashish uchun moʻljallangan.

Shu kabi oʻzbek tili milliy korpusining ogʻzaki nutq korpusini ham ana shunday transkripsiya va teglash orqali yaratishimiz mumkin. Hozirda mavjud ogʻzaki nutq korpuslari yangi materiallar bilan yanada boyitilmoqda, xususan, korpus tarkibidagi audio, multimedia, bolalar nutqi va dialekt korpuslaridagi maʼlumotlarning asosiy qismi ogʻzaki yozib olingan manbalar asosida yaratiladi. Albatta, bu ishni loyihalash juda kata kuch va maʼsuliyatni talab qiladi . Ogʻzaki nutq korpusini loyihalashda bir nechta asosiy bosqichlar va tamoyillarni hisobga olish zarur. Maʼlumotlarni toʻplashda turli manbalarni tanlash kerak. Masalan, oddiy suhbatlar, intervyular. Shuningdek, ishchi nutq yoki jamoat nutqi. Bunda albatta, nutqning kontekstini (muhit, madaniyat, oʻzaro aloqalar) hisobga olish zarur. Va eng asosiy qismi - korpusni transkripsiya qilish. Avvalo, agar ovoqli korpus boʻlsa, nutqni yozma shaklga oʻzgartirish kerak.

Bu jarayonda transkriptni toʻgʻri qilish uchun baʼzi mezonlar mavjud. Jumladan, ovoqli xatoliklar: notogʻri talaffuzlar, jkumlalarni toʻliq yoki qisqacha yozish. Shuningdek, diksiyon va pauzalar: nutqdagi pauzalar, qisqa toʻxtashlar va oʻzgartirishlar ham yozilishiga eʼtibor berish kerak. Markazlashtirilgan transkripsiya: korpusda hamma nutqlar bir xil uslubda yozilishi kerak.

Keyingi bosqich nutq turlarini tahlil qilish va annotatsiya qilishdan iborat. Bunda nutqning talaffuzi va tovushlaridagi oʻzgarishlarni belgilash, nutqdagi grammatik tuzilmalarni tahlil qilish, xususan, sintaktik xatoliklar yoki oddiy nutq elementlarini ajratish, nutqdagi maʼno va mavzularni tahlil qilish va nutqdagi his-tuygʻularni (ijobiy, salbiy, neytral) tahlil qilishdan iborat boʻladi.

Keyingi muhim bosqichlardan biri korpusni saqlash va tahlil qilidirish. Korpusni ma'lum formatda saqlash kerak. Odatda, bu formatlar CSV, JSON, XML, yoki tez tahlil qilish uchun maxsus nutq tahlili formatlarida bo'ladi. Korpusni katta hajmda yig'ish va tahlil qilish uchun ma'lumotlar bazasidan foydalanish mumkin (masalan, SQL, NoSQL). Korpusni tahlil qilish uchun tegishli vositalardan foydalanish zarur. Masalan, lingvistik tahlil qilish uchun NLP (Natural Language Processing) vositalari. Shuningdek, korpusni doimiy ravishda yangilab borish zarur, chunki nutq va til doim o'zgarib turadi, yangi ma'lumotlar va kontekstlar kiritilishi kerak.

Xulosa. Xulosa qilib aytganda, o'g'zaki nutq kompyuter korpuslari zamonaviy tilshunoslik va sun'iy intellekt texnologiyalari rivojida beqiyos ahamiyat kasb etadi. Ular tabiiy tilni chuqur tahlil qilish, nutqni avtomatik tanish va tushunish, shuningdek, interaktiv tizimlar yaratishda asosiy manba bo'lib xizmat qiladi. Ayniqsa, o'zbek tili kabi raqamli resurslari cheklangan tillar uchun bunday korpuslarning yaratilishi va takomillashtirilishi nafaqat ilmiy, balki madaniy va ijtimoiy ahamiyatga ham ega. Shunday ekan, o'g'zaki nutq korpuslarini shakllantirish, ularni boyitish va ulardan samarali foydalanish bugungi kunda dolzarb masalalardan biri hisoblanadi.

Foydalanilgan adabiyotlar :

1. Abdurahmonova N.Z.. Korpus lingvistikasi. GlobeEdit, 2023
2. Mirsaid, A., Bibigul, R., Altinbay, S., & Nilufar, A. (2018). Ontology of grammar rules as example of Noun the Uzbek language and Kazakh languages VI international scientific conference" Modern problems of the applied mathematics and information technology. Al Khorezmiy-2018.
3. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
4. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
5. Tony Robinson, Jereon Fransen, David Pye, Jonathan Foote and Steve Renals WSJAMO: A British English speech corpus for large vocabulary continuous speech recognition
6. Collen Richey. Speech Corpora. 2020
7. <https://www.clarin.eu/resource-families/spoken-corpora>
8. https://web.stanford.edu/dept/linguistics/corpora/material/X_Speech_Corpora.pdf
9. <https://ieeexplore.ieee.org/document/479278>

FARIDA AFRO‘Z SHE‘RIY ASARLARIDA FRAZEOLOGIZM VA DIALEKTIZMLAR

Topildiyeva F.R.

O‘zMU Kompyuter lingvistikasi
va amaliy tilshunoslik kafedrasi katta o‘qituvchisi

Annotatsiya: Ushbu maqolada Farida Afro‘z she‘riy asarlarida ekspressivlikni yuzaga chiqarishda frazeologizm va dialektizmlarning o‘rni, o‘ziga xos lingvomadaniy ko‘rinishlari tadqiq etiladi. Mavzuga oid ilmiy fikr, poetik tahlillar, mulohazalar she‘riy asar namunalari misolida dalillab, tahlil etiladi.

Kalit so‘zlar: etnolingvistika, ekspressivlik, poetik asarlari, leksik frazeologik vositalar, frazeologizm, dialektizm.

Abstract: This article studies the role of phraseology and dialectisms in the manifestation of expressiveness in the poetic works of Farida Afroz, their specific linguocultural manifestations. Scientific ideas, poetic analyses, and reflections on the topic are substantiated and analyzed by using the example of her poetic works.

Keywords: ethnolinguistics, expressiveness, poetic works, lexical phraseological means, phraseology, dialectism.

Аннотация: В статье рассматривается роль фразеологии и диалектизмов в выражении экспрессивности в поэтических произведениях Фариды Афроз, а также их специфические лингвокультурологические проявления. На примере поэтических произведений обосновываются и анализируются научные идеи, поэтические анализы и размышления по теме.

Ключевые слова: этнолингвистика, экспрессивность, поэтические произведения, лексико-фразеологические средства, фразеологизм, диалектизм.

Bugungi kun o‘zbek she‘riyatida Farida Afro‘z o‘z uslubiga ega bo‘lgan iste’dodli, o‘zining jo‘shqin va sarmazmun ijodi bilan o‘zbek xalqi qalbidan munosib o‘rin olgan shoiradir. Uning ijod bo‘stonida o‘ziga xos tarzda bitilgan she‘riy asarlari bor bo‘lib, o‘zbek xotin-qizlarining o‘y-fikrlari, orzu-havaslar, baxt-u saodati, kezi kelganda alam-nafrati betakror falsafiy fikrlar bilan yo‘g‘rilganligi, ramziy tashbehlari, shartli sinonimlar, omonimlar, frazeologizmlar va ayrim o‘rinlarda dialektizmlardan, leksik uslubiyatdan o‘rinli foydalana olishi shoiraning she‘rlari til xususiyatini rang-barangligidan va uning so‘z qo‘llash mahorati jilosidan dalolat beradi. Umuman olganda, Farida Afro‘z she‘riyatda dunyo go‘zalligini, hayot falsafasini, ma‘no mazmunini, o‘zini menida, o‘z his-tuyg‘ularini esa vafo, sadoqat, sabr, alam, nafrat, nasihat, ibrat kabilar ko‘rinishida ifoda etadi.

Farida Afro‘zning shunday she‘rlari borki, o‘ziga xos til uslubi bilan, so‘zlarni o‘z ijod gulshanida “ipga tizilgan marjonday” benuqson qo‘llashi bilan, ayniqsa, frazeologik birliklarni va ayrim o‘rinlarda etnolingvistikaga e‘tibor qaratib, dialektizmlarni ham o‘rinli ifodalashi bilan, yana har qanday kitobxonni dardini kuylagandek his ettirib, misralarda fikrning indallosi ifodasini baralla aytib, go‘yo misralar og‘ushiga chorlagandek bo‘ladi.

Uning yozganlari go‘yo har qanday iztiroblarga qarshi tura olishda, ba‘zan ro‘baro‘ kelganda kishiga kuch, dalda, malham bo‘la oladi. Yana shoira she‘riyatida kezi kelganda nasihat, ba‘zan nafrat yaqqol ko‘zga tashlansa-da, samimiylik ufurib turadi misralarni uyg‘unligida nafaqat ma‘nodosh so‘zlar, takrorlar, balki, frazeologizm va dialektizmlar ham marjonday ipga tizilganini ko‘ramiz.

Masalan, “Ovvoramiz” sarlavhali she‘rida:

Har kishiki **ko‘zi ochdir**,

Ko‘r qilar, to‘zoni ganj,

Yo'l ko'rsatar sarbon qolib,
Karvon bilan ovvoramiz,- deydi shoira.

Ushbu misrada Farida Afro'z "ko'zi och" frazeologizmini o'rinli qo'llab ta'sirchanlik doirasini go'yo tahlilini chizib berganday ifodalagan. Xalqimiz nazdida ko'zi ochlik uzoq davrlardan ma'lum bo'lgan qusurdir. Ma'lumki, boylik ilinji, hatto boylik to'zoni ham ko'zi och, tor doirani his qilgan har qanday insonni ko'r qilishi tabiiy hol ekanligini bilamiz. Bu borada shoira yo'l boshlovchi sarbonni unutib, karvonni markazga qo'yib uni bilan ovvoralikni, sargashtalikni shoirona tarzda ifodalagan. To'zon so'zi izohli lug'atda aslida havoda ko'tarilgan mayda narsalar, xas-xushlar, chang ma'nosida qo'llanadi. Farida Afro'z bo'lsa to'zoni ganj deya boylik xabarini nazarda tutgan.

Masalan: O'choq boshida qumg'on sharaqlab qaynab, cho'g' ustiga to'kildi, bug' va kul to'zoni ko'tarildi. A.Muxtor. "Opa –singillar" asari.

Farida Afro'z esa lingvokulturologik munosabatni "to'zoni ganj" ifodasi bilan ifodalab, boylik shamoli, gardi, xabari, sasi ma'nosida qo'llagan.

Yana "Kechirmoqni bilsang agar" sarlavhali she'rida:

Og'zing oshga yetganda voh,
Kimdir tuproq sochar pinhon,
Senga ravo bo'lmagan o',
Taomlardan qutulgayсан,- deydi shoira.

Shoira she'riyatida tilshunoslik va adabiyotshunoslikning uyg'unligi ko'zga yaqqol tashlansa-da, ifodalarning til xususiyatlari o'ziga xos bo'lib "og'zing oshga yetganda" frazeologizmi orqali estetik ta'sirni yanada kuchaytiradi, tinglovchi buni albatta his qiladi.

Yoki yana bir she'rida:

Sog'ina, sog'ina sog'inmay qo'ydim,
Endi menga baribir kelsang, kelmasang
Qo'lingdan kelsa gar, **osmonni tashla**
Endi menga baribir sevsang, sevmasang.

Ushbu misrada esa shoira "osmonni tashlamoq" frazeologizmidan o'rinli foydalanadi.

Qo'lingdan kelsa gar osmonni tashla, bu ifoda orqali o'ziga ishonch bildiradi, hatto osmonni tashlab yuborsa ham endi baribir, kelsang, kelmasang, sevsang, sevmasang baribir, ahamiyati yo'q deya baralla ayta oladi. O'z irodasiga ishongan inson talqinida ekspressivlik namoyon bo'ladi. Ushbu o'rinda osmonni tashla frazeologizmi nutqiy ta'sirchanlikni oshiradi, tinglovchini diqqatini tortadi.

Yana:

Sog'ina, sog'ina sog'inmay qo'ydim,
Xo'p telba ko'ngilga cho'pchaklar aytdim,
Aldashni o'rgandim yurakni boplab,
Daryoga oborib sug'ormay qaytdim,-deydi shoira.

Ushbu misrada umidlarini mosuvo qilish, chippakka chiqarish kabi asnoda ta'riflar berilgan. Yana mastona telba ko'ngilga hatto cho'pchaklar aytib, hatto yurakni boplab aldashni o'rganganiga e'tibor bersak, lingvomadaniy ta'sir doirasini Afro'zona uslubidan darak beradi va "Daryoga oborib sug'ormay qaytdim" deya baralla aytadi. Frazeologik birlik ma'nosi oydinlashganda ifodaning qanchalik ta'sirchan ekanligi, metaforik tafakkurning qanchalik ravon va boy ekanligidan dalolatdir. Uning she'riyatida so'zlarning dialektizm va frazeologizmlarning ekspressivlik doirasida qo'llanishi ham janrlar kabi rang-baranglik tasnifini ifodalaydi desak mubolag'a bo'lmaydi. Aksincha, noan'anaviylik yagona, etnolingvistik, lingvomadaniy so'z qo'llash mahoratidan darakdir.

Shoira yana bir sherida shunday deydi:
Seni xayolimdan yashirolmayman
Uzib otolmayman seni sug'urib
Sen-la ketolmayman
She'rlar bitolmayman
Zaqqumlar yutib.

Zaqqumlar yutib iborasini qo'llaydi, ya'ni xayolidan ham yashirolmagan alam va dardni ifodasini beradi. Zaqqum yutmoq – zahar yutmoq, achchiq alam yutmoq, azob va izzat chekmoq ma'nolarida qo'llanadi. Shoira yorning ozorlarini, undan hatto xayolida ham kecholmasligini zaqqumlar yutmoq frazeologizmi orqali oydinlashtiradi.

Farida Afro'z she'riyatida hayrat, taajjub, ehtiros, nafrat, sevinch, surur, o'kinch, yig'i, sabr va alam, armon va qanoat, yana g'urur ufurib turadi. Shoira turfa his –tuyg'ular silsilasida yashaydi, ma'naviy intellektual ijtimoiy faoliyatida o'zini hayajonga solgan tuyg'ularini o'ziga xos lingvokulturologik ifodalar bilan boyitadi. Masalan:

Til tegizma, millatimga, qo'y,
Daho tuqqan har yili onam,
Otam **faq-a-a-t** poda boqmagan.

Ushbu misrada "til tegizmoq" frazeologizmini qo'llaydi, yana Farida Afro'z so'zning fonetik qiyofasiga ham "faq-a-a-t" alohida e'tibor qaratadi.

Yana bir misol:

Axir seni ko'rganim tushmas,
Xayolingsiz hushlarim qushmas,
Osmon sinib, zaminga tushmas,
Kel, bugunni bayram qilaylik, - deydi shoira.

Osmon uzilib, yerga tushmas frazeologizmini, osmon sinib, zaminga tushmas shaklida ifodalaydi va o'zining lingvomadaniy ifodasi bilan kayfiyat ulashadi.

Kechirmoqni bilsang agar she'rda:

Tog'day og'ir, **zilu zambil**
Qasos ezmas yelkalarim,
O'ch- intiqom o'ti yongan
Alamlardan **qutulgaysan.**

Yoki ikki tilli ilon yanglig' odamlardan qutulgaysan, - deydi

Shoiraning ushbu etnolingvistik, dialektal so'zlarni she'riyatida qo'llashi lingvomadaniy xususiyatlarni soddalik bilan tinglovchiga yetkazishdir, masalan: ko'zimga suray, tutolgin, oborib, yarodir, yuvolmas, bosingan, tozarmoq, o'tolmaysan, yo'g'ingni, huvildoq, izzillab, unutmog', suyunib kabilarni misralarda o'ziga xos ohangda keltirgan..

Xulosa qilib aytadigan bo'lsak, Farida Afro'z she'riy asarlari bugungi kun o'zbek she'riyatida lingvokulturologik tahlilga asos bo'la oladigan manba vazifasini bajara oladi. U she'riy asarlarida frazeologizm va dialektizmlarni Afro'zona uslubda ifodalashi bilan til xususiyatlarini yanada o'ziga xos talqini asosida ohangdorlikni, ekspressivlikni jilolantiradi. Bu esa Farida Afro'z she'riyatining, intim lirika sifatida betakrorligidan, so'zlarni tanlab, ma'no xususiyatlarini ifodalashda mentallik nishonasidan darak beradi.

Foydalanilgan adabiyotlar ro'yxati:

1. А. Абдуллаев. Ўзбек тилида экспрессивлик ифодалашнинг синтактик усули. Тошкент. 1987.
2. Афрўз Фарида (Бўтаева). "Ўша кун - бугундир". Сайланма. Тошкент:

3. “Шарқ” 2013.
4. O‘ZBEK TILINING IZOHLI LUG‘ATI. BESHINCHI JILD. G‘AFUR G‘ULOM.TOSHKENT- 2022.
5. Farida Afro‘z. Tashbeh. T., “Sharq”, 2007.
6. Farida Afro‘z.Ko‘zim manim.-T.G‘afur G‘ulom nomidagi Adabiyot va san’at nashriyoti.1999.
7. Ёўлдошев М. Бадий матн ва унинг лингвопоэтик таҳлили асослари. Тошкент: Фан ,2007.

NER TEXNOLOGIYASINING MATN TAHLILIDAGI AHAMIYATI

Xoljurayeva Yulduz Sobir qizi

O'zMU Kompyuter lingvistikasi yo'nalishi
1-bosqich magistranti

E-mail: abdirahimovayulduz4@gmail.com

ORCID-0009-0008-8120-8239

Umirzaqova Maxina Anvarjon qizi

O'zMU Kompyuter lingvistikasi yo'nalishi
1-bosqich magistranti

E-mail: mahinasabirzanov@gmail.com

Orcid 0009-0001-6355-7645

Annotatsiya, mazkur maqolada nomlangan entitylarni aniqlash (NER – Named Entity Recognition) texnologiyasining nazariy asoslari, amaliy qo'llanilishi va o'zbek tili uchun moslashtirish zarurati keng yoritilgan. Maqolada NER texnologiyasining turli amaliy sohalaridagi — yangiliklar monitoringi, savol-javob tizimlari, ijtimoiy tarmoqlar tahlili va sohaviy (tibbiy, huquqiy) matnlar bilan ishlashdagi afzalliklari misollar bilan ko'rsatib berilgan. Ayniqsa, O'zbek tilining morfologik murakkabligi va so'z tartibidagi erkinlik bu sohada maxsus moslashtirilgan, chuqur o'rganishga asoslangan yondashuvlarni talab qilishi qayd etiladi.

Kalit so'zlar: Tabiiy tilni qayta ishlash (NLP), chuqur o'rganish, ichma-ich entitetlar, ketma-ketlikni belgilash, morfologik murakkablik, axborotni ajratib olish.

Annotation, this article explores the theoretical foundations, practical applications, and the necessity of adapting Named Entity Recognition (NER) technology for the Uzbek language. It highlights the diverse use cases of NER in areas such as news monitoring, question-answering systems, social media analysis, and domain-specific (medical, legal) text processing, illustrated through relevant examples. The article particularly emphasizes that the morphological complexity and flexible word order of the Uzbek language require specially adapted, deep learning-based approaches in this field.

Keywords: Natural language processing (NLP), deep learning, nested entities, sequence labeling, morphological complexity, information extraction.

Аннотация, В данной статье подробно освещены теоретические основы технологии распознавания именованных существностей (NER – Named Entity Recognition), её практическое применение и необходимость адаптации для узбекского языка. В статье с примерами показаны преимущества применения NER в различных прикладных областях — мониторинге новостей, системах вопросов и ответов, анализе социальных сетей, а также при работе с специализированными текстами (медицинскими, юридическими). Особое внимание уделено тому, что морфологическая сложность узбекского языка и свобода в порядке слов требуют специально адаптированных, основанных на глубоком обучении подходов.

Ключевые слова: Обработка естественного языка (NLP), глубокое обучение, вложенные сущности, разметка последовательностей, морфологическая сложность, извлечение информации.

Nomlangan entitylarni aniqlash (NER – Named Entity Recognition) – bu berilgan tuzilmasiz matnda ma'lum sinflarga (masalan, Shaxs, Tashkilot, Joy va hokazo) tegishli obyektlarni aniqlashga qaratilgan vazifadir. Bu vazifaning amaliy qo'llanilish sohalariga yangiliklar va ijtimoiy tarmoqlar tahlili, kontent tavsia qilish tizimlari, hamda tibbiy matnlardan

ma'lumot ajratib olish kiradi. Har xil qo'llanilish holatlari turli sinflardagi entitylarni talab qiladi. Ko'pincha nomlangan entitylar otlarga to'g'ri keladi, lekin ular umumiy otlar yoki sonlar (telefon raqamlari, miqdorlar, sanalar) ham bo'lishi mumkin.

NERning kichik vazifalaridan biri bu ichma-ich nomlangan entitylarni aniqlash (NNE – Nested Named Entity recognition) hisoblanadi. Uning asosiy farqi shundaki, bunda bir entity boshqa, yirikroq entity tarkibida joylashgan bo'lishi mumkin. Masalan, “Московский государственный университет имени М.В.Ломоносова” matnida ikki entity mavjud: butun ibora tashkilot sifatida, “М.В.Ломоносова” esa shaxs nomi sifatida tan olinadi. NNEga zid bo'lgan holat bu soddalashtirilgan (yassi) entitylar bo'lib, ularning ichida boshqa obyektlarni NER algoritmlari aniqlay olmaydi. Hozirgi mavjud NER yechimlarining aksariyati aynan soddalashtirilgan entitylar bilan ishlaydi, holbuki real matnlarda ichma-ich entitylar ko'proq uchraydi. Ayni paytda soddalashtirilgan NER uchun ishlatiladigan metodlarning ko'pi bu vazifani ketma-ketlikni belgilash (sequence labeling) vazifasi sifatida bajaradi. Shu bilan birga, bu ketma-ketlik asosidagi metodlar ichma-ich entitylarni ham aniqlashda ishlatilishi mumkin, xuddi giper-graflarga asoslangan metodlar kabi.¹¹⁴

Nomlangan entitetlar (Named Entities) — bu shaxslar, tashkilotlar va joy nomlarini o'z ichiga olgan iboralardir. Masalan: Aziz Toshkentga ketdi.

Ushbu gapda ikkita nomlangan entitet mavjud: Aziz — shaxs, Toshkent — joy nomi.

Nomlangan entitetlarni aniqlash (NER) — axborotni ajratib olish tizimlarining muhim vazifasidir. Ayniqsa ingliz tili uchun bu sohada juda ko'p tadqiqotlar olib borilgan.¹¹⁵ Message Understanding Conferences (MUC) tadqiqotchilarga ingliz tili uchun bir xil ma'lumotlarda tizimlarini baholash imkonini berdi va entitetlarni belgilash sxemasini ishlab chiqdi.¹¹⁶ CoNLL-2003 umumiy vazifasi esa tilga bog'lanmagan nomlangan entitetlarni aniqlashga qaratilgan. Bu tadqiqotda quyidagi to'rtta entitet turi asosiy e'tiborga olingan:

Shaxslar (PER)

Joy nomlari (LOC)

Tashkilotlar (ORG)

Boshqa — oldingi toifalarga kirmaydigan entitetlar (MISC) (masalan: "Italian" yoki "1000 Lakes Rally").

Nomlangan entitetlarni aniqlash (NER – Named Entity Recognition) tabiiy tilni qayta ishlashdagi muhim vazifalardan biri bo'lib, matndagi entitetlarga ishoralarni aniqlashga qaratilgan. Bu vazifa birinchi marta Message Understanding Conference (MUC-6) da qisqa muddatli topshiriq sifatida taqdim etilgan.¹¹⁷ O'sha paytda bu vazifani nisbatan qisqa vaqt ichida amalga oshirsa bo'ladi deb o'ylangan, shuningdek, bu boshqa axborotni ajratib olish vazifalari uchun domenlarga bog'liq bo'lmagan vosita sifatida xizmat qilishi mumkin deb hisoblangan.

NER vazifasini bajarish uchun matndan ma'lum qismlar ajratib olingan, shunda bu qismlar har qanday kontekstdagi xuddi shu entitetga ishora qiladi va mavjud bo'lmagan kontekstlarda boshqa narsaga ishora qilmaydi – ya'ni ular qat'iy belgilovchilardir.¹¹⁸

Quyidagi misollar ushbu vazifaning murakkabligini ko'rsatadi. Birinchidan, bitta entitet turli birikmalar orqali ifodalanishi mumkin: “JFK”, “Kennedy” yoki “John F. Kennedy” iboralari kontekstga qarab AQShning 35-prezidentiga ishora qilishi mumkin. Boshqa holatda esa, “JFK” yoki “John F. Kennedy” iboralari Nyu-Yorkdagi aeroportni bildirishi mumkin. MUC-6 ta'rifida eng muhim belgilovchilar sifatida shaxs nomlari, geografik joylar va tashkilot nomlari ko'rsatilgan bo'lib, bu NER vazifasining eng oddiy ta'rifidir.

¹¹⁴ Abrosimov K.I., Mosyagina A.G. Sodner for Russian nested named entity recognition // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue". 2002. Issue 21. P.2.

¹¹⁵ Andrew Borthwick. 1999. A Maximum Entropy Approach to Named Entity Recognition. PhD thesis, New York University. P. 22.

¹¹⁶ Nancy Chinchor, Erica Brown, Lisa Ferro, and Patty Robinson. 1999. 1999 Named Entity Recognition Task Definition. MITRE and SAIC. P.2.

¹¹⁷ Grishman, R., and Sundheim, B. 1996. Message understanding conference-6: a brief history. In Proceedings of the 16th Conference on Association for Computational Linguistics, pp. 466–71.

¹¹⁸ Kripke, S. 1982. Naming and Necessity. Boston: Harvard University Press. P. 4-5.

Ushbu texnologiya tuzilmagan matnlarni strukturalashtirish, kontekstual tahlil qilish hamda axborotni samarali indekslashda asosiy vosita bo'lib xizmat qiladi.

NER texnologiyasining ahamiyatini quyidagi amaliy holatlar misolida ko'rish mumkin:

1. Tuzilmagan matnni tuzilgan shaklga keltirish. Ya'ni ko'plab matnlar, xususan yangiliklar, rasmiy xabarlar va bloglar tuzilmagan holatda bo'ladi. Masalan: "Dildora 2025-yil 25-aprel kuni Toshkent shahriga konsertga bordi."

NER tizimi ushbu jumladan "Dildora" (shaxs), "2025-yil 25-aprel" (sana), "Toshkent" (joy) kabi entitetlarni ajratib oladi. Bu esa avtomatik tasniflash va axborotni tuzilgan shaklga keltirishga zamin yaratadi.

2. Maqsadli qidiruv va savol-javob tizimlarida qo'llanilishi. NER texnologiyasi foydalanuvchi savolidagi entitetlarni aniqlash orqali kontekstga mos javob qaytarilishini ta'minlaydi. Masalan: "Amir Temur qachon tug'ilgan?" Bu yerda "Amir Temur" shaxs sifatida aniqlanadi va tizim mazkur entitet asosida mos ma'lumotni qidiradi.

3. Ijtimoiy tarmoqlar va yangiliklar monitoringi

NER texnologiyasi ijtimoiy tarmoq postlari, yangiliklar va foydalanuvchi sharhlarida tilga olinayotgan joylar, shaxslar yoki tashkilotlarni aniqlash orqali voqealarning kontekstini o'rganishga yordam beradi. Masalan: "Toshkent shahrida Coca-Cola tomonidan tashkil etilgan sport musobaqasi katta qiziqish uyg'otdi." Bu jumladan "Toshkent" (joy), "Coca-Cola" (tashkilot), "sport musobaqasi" (voqea) kabi entitetlar ajratiladi.

4. Soha bo'yicha ixtisoslashgan tahlillar

NER texnologiyasi tibbiyot, huquq, moliya kabi sohalarda aniq entitetlarni ajratish orqali ilmiy tahlil va statistik hisob-kitoblarni osonlashtiradi. Masalan: "Bemorga Paratsetamol va Amoksitsillin dori vositalari tavsiya etildi." NER ushbu matndan "Paratsetamol" va "Amoksitsillin" dori vositalarini dori nomlari sifatida ajratib oladi.

5. O'zbek tili uchun maxsus NER modellari yaratish zaruriyati

O'zbek tili erkin so'z tartibi va boy morfologik tizimga ega bo'lgani sababli, umumiy NER modellari O'zbek tilida yetarlicha yuqori aniqlikda ishlamaydi. Masalan, "Andijon" so'zi shahar nomi ham, inson ismi ham bo'lishi mumkin. Bundan tashqari, affiksial o'zgarishlar — "Toshkentda", "Toshkentga", "Toshkentlik" kabi shakllar — entitetlar klassifikatsiyasini murakkablashtiradi. Shu sababli, O'zbek tili uchun maxsus korpuslar, qo'lda belgilangan (annotatsiyalangan) ma'lumotlar bazasi va chuqur o'rganish asosidagi yondashuvlar muhim hisoblanadi

Foydalanilgan adabiyotlar:

1. Abdurakhmonova, N., Barakhnin, V., Mengliev, D., & Eshkulov, M. (2024). Developing named entity recognition algorithms for Uzbek: Dataset insights and implementation. Data in Brief.
2. Abdurakhmonova, N., Mengliev, D., Barakhnin, V., Eshkulov, M., & Palvanov, B. (2023). Dictionary-based medical text analysis in Uzbek: overcoming the low-resource challenge. Proceedings of the 2023 IEEE Ural-Siberian Conference on Computational Technologies (USCCT).
3. Abrosimov K.I., Mosyagina A.G. Sodner for Russian nested named entity recognition // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue". 2002. Issue 21.
4. Andrew Borthwick. 1999. A Maximum Entropy Approach to Named Entity Recognition. PhD thesis, New York University.
5. Grishman, R., and Sundheim, B. 1996. Message understanding conference-6: a brief history. In Proceedings of the 16th Conference on Association for Computational Linguistics, pp. 466–71.
6. Kripke, S. 1982. Naming and Necessity. Boston: Harvard University Press.
7. Turian, J., Ratnoff, L., & Bengio, Y. (2010). "Word Representations: A Simple and General Method for Semi-Supervised Learning."

SAVODXONLIKNI OSHIRISHDA MATNLARNI AVTOMATIK TAHRIR QILISH TIZIMINING O'RNI

Askarova Umidaxon Maxamadjon qizi
O'zbekiston Milliy universiteti o'qituvchisi
E-mail: samina090822@gmail.com

Annotatsiya: Dunyoda Kompyuter lingvistikasiga doir qator izlanishlar olib borilmoqda, jumladan, mashina tarjimasi, korpus texnologiyasi, kompyuter leksikografiyasi hamda dasturiy ilovalarning lingvistik asoslarini ishlab chiqish, leksikografik matnlarga ishlov berish va lingvostatistik tahlil etish borasida muayyan darajada natijalarga erishib kelinmoqda. Axborot almashinuvining jadallashuvi va ma'lumotlar ko'lamining tobora kengayib borayotgan bir paytda bilvosita tahrirga bo'lgan talab tabiiyki oshib bormoqda. Biz ham ushbu tadqiqot ishimizda avtomatik tahrir va uning savodxonlikni oshirishdagi ahamiyati haqida batafsil to'xtalib o'tganmiz.

Kalit so'zlar: kompyuter, kompyuter lingvistikasi, texnologiya, matn, avtomatik tahrir, imlo, so'z, kontekst.

Аннотация: В мире проводится ряд исследований по компьютерной лингвистике, в том числе достигаются определенные результаты в области машинного перевода, корпусной технологии, компьютерной лексикографии и разработки лингвистических основ программных приложений, обработки лексикографических текстов и лингвостатистического анализа. В условиях интенсификации обмена информацией и расширения объема данных, естественно, растет спрос на косвенное редактирование. В этой исследовательской работе мы также подробно обсудили автоматическое редактирование и его значение в повышении грамотности.

Ключевые слова: компьютер, компьютерная лингвистика, технология, текст, автоматическое редактирование, орфография, слово, контекст.

Abstract: In the world, a number of studies are being conducted in the field of computational linguistics, including machine translation, corpus technology, computer lexicography, as well as the development of the linguistic foundations of software applications, the processing of lexicographic texts, and linguostatistical analysis. At a time when the exchange of information is accelerating and the volume of information is constantly expanding, the demand for indirect editing is naturally growing. In this research work, we also discussed automatic editing in detail and its importance in improving literacy.

Keywords: computer, computational linguistics, technology, text, automatic editing, spelling, word, context.

Bugungi kunga kelib hayotimizni kompyuter texnologiyalarisiz tasavvur qila olmaymiz. Bu elektron texnologiya har bir sohaning ajralmas qismiga aylanib ulgurgan deya baralla ayta olamiz. Yildan yilga har bir soha rivojlanayotganidek, kompyuter dasturlari bilan bog'liq yangi yo'nalishlari yuzaga kelmoqda. Buning yaqqol misoli sifatida kompyuter lingvistikasining asosiy yo'nalishi sifatida namoyon bo'lgan avtomatik tahrir va tahlilni keltirishimiz mumkin. Avtomatik tahrir yo'nalishi XX asr o'rtalariga kelib rivojlandi va so'ngi yillarda ko'plab mutaxassislar tomonidan yanada rivojlantirilmoqda. Hozirgi kunda barcha kompyuterlarda avtomatik tahrirlash dasturlari bor. Buning yaqqol misoli sifatida har bir kompyuter qurilmasida mavjud bo'lgan MS DOS operatsion tizimiga joylashgan rus va ingliz tillaridagi matnlarni avtomatik tarjima qiluvchi dasturni keltirishimiz mumkin. Ular o'zbekcha matnlardan tashqari ruscha, inglizcha, nemischa kabi bir qancha tillardagi matnlarni ham tahrirlaydi. Ushbu yo'nalishda R.G.Kotov, V.G.Britvin,

I.A.Melchuk, A.B.Kuznetsov, L.I.Belyayeva, V.A.Chijakovskiy, G.G.Belogonov, I.S.Duganovlar samarali tadqiqot olib borgan.

Bundan tashqari internet tizimida mavjud tilmoch veb sahifasidagi “Tahrirchi” bo‘limi yoki matn.uz veb-sahifalari ham matnni avtomatik tahrir qilishda juda keng qo‘llanib kelinmoqda. Ushbu veb sahifalar o‘zbek tilida lotin va kirill grafikasida bexato yozishda beminnat ko‘makchi vazifasini bajarmoqda desak mubolag‘a bo‘lmaydi.

Biz bilamizki, kompyuter yaratilmasdan avval matnlar yozuv mashinkalarida terilgan. Bu jarayonda ko‘plab xatolarga yo‘l qo‘yilgan. Misol uchun: grammatik, punktuatsion, orfografik, stilistik. Ushbu holatda matnni boshdan oxirigacha qayta yozib chiqishga to‘g‘ri kelgan, buning natijasida esa foydalanuvchiga ham vaqt tomondan va unga sarflanadigan kuch tomondan ham bir qancha qiyinchilik tug‘dirgan. Avtomatik tahrirlash jarayonida MS Word lug‘atida mavjud bo‘lmagan so‘z uchrasa, ostiga qizil chiziq bilan ajratib ko‘rsatiladi. Mazkur holatda dastur so‘z imlosi haqida variant taklif eta olmaydi, chunki MS Word lug‘atida bu so‘z mavjud emas. Ehtimol u so‘z to‘g‘ri yozilgan bo‘lishi mumkin, bunda insonning ishtiroki zarur bo‘ladi. Agar so‘zning imlosi to‘g‘ri bo‘lsa, matnda tez-tez qayta ishlatilsa, uning avtokorreksiya to‘g‘ri deb sanalishi uchun MS Word lug‘atiga “kiritib qo‘yish” (kontekst menyudagi “add” yoki “добавить” komandasi yordamida) mumkin. Matn kiritish davomida so‘z yoki so‘zlar birikmasi ko‘k chiziq bilan ajratib qo‘yiladi, bunda so‘zning imlosida xato mavjud bo‘ladi. Bu holda foydalanuvchi xatoni o‘zi to‘g‘rilab qo‘yishi zarur yoki ostiga chizilgan so‘zda sichqonchani o‘ng tomoni bosilsa, kontekst menyusida so‘zning to‘g‘ri varianti taklif etiladi va uni tanlashi mumkin. [Xolmanova, 2019] Yuqorida keltirilgan fikrlardan kelib chiqib, demak, kompyuterda matnni tahrirlash birmuncha qulayligini bilishimiz mumkin.

Ammo, tanganing ikki tarafi bor deganlaridek, yuqorida keltirilgan ushbu qulayliklar ba’zi foydalanuvchilar uchun salbiy oqibatlariga olib kelmaydimi degan fikr ham tug‘iladi insonda. Chunki, biz universitetda o‘quv mashg‘ulotlarda talabalardan bexato yozishni so‘raganimizda ayrim talabalarning “Imloviy yoki uslubiy to‘g‘ri yozishimiz shart emas, kompyuterda matnni to‘g‘rilab beradigan programmalar bor” deyishlarini eshitib, achinasan kishi. Avtomatik tahrir va tahlil jarayoni makro matnlarni vaqtdan unumli foydalangan holda to‘g‘rilashdir. Biz bir narsani unutmasligimiz kerakki, biz kompyuterdan qanchalik oqilona foydalanishimiz albatta savodxonligimizga zarar keltirmasligi kerak. Kompyuterdan oqilona foydalanish yoki unga qaram bo‘lish ham insonning o‘ziga bog‘liq. Internetni pichoqqa o‘xshatilganini juda ko‘p eshitganmiz. Kompyuterni ham yuqorida keltirib o‘tilgan ijobiy va salbiy oqibatlaridan kelib chiqib pichoqqa qiyoslash mumkin, nazarimda. Chunki, xuddi pichoqdek undan oqilona foydalansak, bizga foyda beradi. Aks holda zarar berishi muqarrardir.

Xulosa qilib shuni aytish mumkinki, zamonaviy texnologiyalarning kundalik turmushda muhim o‘rin egallashi, ilm-fanning rivojlanishi avtomatik tahrir va tahlilning ahamiyatini oshirmoqda. Avtomatik tahrir faqat matnlardagi mexanik xatolarni bartaraf etibgina qolmay, matnning savodli yozilishini ham ta’minlaydi, matn kirituvchiga jumlada xato so‘z borligi haqida xabar berib, xatoni bartaraf etishga undaydi, natijada to‘liq lingvistik me’yorlar darajasidagi matnni yuzaga keltirishda vaqt tejiladi.

Foydalanilgan adabiyotlar:

1. Abdurakhmonova N., Tuliyeu U., Ismailov, A., Abduvahobov G. Uzbek electronic corpus as a tool for linguistic analysis. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240). 2022.
2. Abdurakhmonova N., Abduvahobov G. O‘quv lug‘atini tuzishning nazariy metodologik asoslari. Международный журнал искусство слова, 4(6). 2021.
3. Askarova U. Kompyuter lingvistikasining kirib kelishi va uning bugungi kundagi ahamiyati. American journal of education and learning (pp. 538-543). 2025, February.

4. Raxmanov X.E., Sunatov J.T. O‘zbek tili kompyuter lingvistikasi yo‘nalishida olib borilgan ilmiy tadqiqotlar, maqola “Kompyuter lingvistikasi: muammolar, yechim, istiqbollar” Xalqaro ilmiy-amaliy konferensiya, Toshkent, 2022.
5. Sobirova N. Avtomatik tahrir asosiy yo‘nalish sifatida. Academic Research in Educational Sciences. Multidisciplinary Scientific journal (pp. 77-82). 2024, June.
6. Xolmonova Z.T. Kompyuter lingvistikasi o‘quv qo‘llanma, Toshkent, 2019

JAHON TILSHUNOSLIGIDA OMONIMLARNING KORPUSGA ASOSLANGAN LUG‘AT BAZASINI YARATISH MASALASI

Anvarova Durdona

PhD student, National University of Uzbekistan

E-mail: durdonanvarova167@gmail.com

Annotatsiya: Maqolada jahon tilshunosligida omonimlarning korpus lingvistikasi asosida chuqur tadqiq etilishi va ularning avtomatlashtirilgan lug‘at bazalarini shakllantirish masalasi keng yoritiladi. Zamonaviy kompyuter lingvistikasi, sun‘iy intellekt yondashuvlari va mashinaviy o‘rganish asosida omonimlarni aniqlash, ularni kontekstual farqlash va semantik guruhlariga ajratish tamoyillari ko‘rib chiqiladi. Buyuk Britaniya, AQSh, Germaniya, Rossiya, Xitoy kabi davlatlar tajribasi asosida korpusga asoslangan modellar orqali omonimlarning tilda real qo‘llanilishi, chastotasi va grammatik xususiyatlari o‘rganiladi. Tadqiqot natijalari leksikografiya, tarjima, NLP, til o‘rgatish va avtomatik semantik analiz sohalarida muhim amaliy ahamiyatga ega.

Kalit so‘zlar: korpus lingvistikasi, omonim, semantik tafovut, leksik baza, sun‘iy intellekt, so‘z ma‘nosini ajratish, WSD, kollokatsiyalar, WordNet, COCA, BNC, DeReKo

Annotation: This article explores the development of automated homonym lexicon databases based on corpus linguistics in global linguistics. It examines methods of identifying homonyms, distinguishing their meanings in context, and classifying them into semantic categories using modern computational linguistics, artificial intelligence, and machine learning. Based on the experience of countries such as the UK, USA, Germany, Russia, and China, the study investigates the actual usage, frequency, and grammatical features of homonyms through corpus-based models. The findings have significant practical implications for lexicography, translation, NLP, language teaching, and automatic semantic analysis.

Keywords: corpus linguistics, homonym, semantic distinction, lexical database, artificial intelligence, word sense disambiguation, collocations, WordNet, COCA, BNC, DeReKo

Аннотация: В статье рассматривается проблема создания автоматизированных лексикографических баз данных омонимов на основе корпусной лингвистики в мировой лингвистике. Освещаются методы выявления омонимов, их контекстного различения и семантической классификации с использованием современных подходов компьютерной лингвистики, искусственного интеллекта и машинного обучения. На основе опыта Великобритании, США, Германии, России и Китая анализируется реальное употребление, частотность и грамматические особенности омонимов с использованием корпусных моделей. Результаты исследования имеют важное практическое значение для лексикографии, перевода, НЛП, преподавания языков и автоматического семантического анализа.

Ключевые слова: корпусная лингвистика, омоним, семантическое различие, лексическая база, искусственный интеллект, дисамбигуация значений, коллокации, WordNet, COCA, BNC, DeReKo

Tilshunoslikning zamonaviy bosqichida leksik birliklarning ko‘p qirrali semantik xususiyatlarini aniqlash va ularni tizimlashtirish muhim masalaga aylanmoqda. Omonimlar – bu fonetik va morfologik shakli bir xil, ammo semantik jihatdan har xil so‘zlar bo‘lib, kompyuter tilshunosligi, avtomatik tarjima, sun‘iy intellekt tizimlari uchun jiddiy sinov hisoblanadi [Navigli, 2009]. Ularni avtomatik tarzda ajratish uchun zamonaviy korpus lingvistikasi yondashuvlari muhim ahamiyatga ega.

Omonimlarning tabiati ko‘p qirrali bo‘lib, ular polisemiyaga yaqinlik jihatidan ham leksik-semantik murakkablik tug‘diradi. Ular orasidagi asosiy farq – omonimlar mutlaqo alohida

semantik yadrolarga ega bo'lsa, polisemiyalar bitta semantik ildizdan kelib chiqqan turli ma'no variantlarini bildiradi. Masalan:

- "Oq" — (rang) va "oq" — (oqmoq) – morfologik o'xshashlik, ammo semantik yadro butunlay farqli.
- "Yoz" — (fasl) va "yoz" — (yozish amali) — kontekstga bog'liq holda farqlanadi.

Shu jihatdan, ularni kompyuter tizimlari tomonidan to'g'ri farqlash uchun kontekstual modellarga asoslangan yondashuvlar zarur. Jahon tilshunosligida bugungi kunda omonimlarning tarjimai bilan bog'liq muammolar korpusga asoslangan tarjima yutuqlari orqali hal qilinadi.

Korpus lingvistikasi – tilni real qo'llanilishda o'rganishga asoslangan ilmiy yondashuv bo'lib, omonimlarni tahlil qilishda quyidagi imkoniyatlarni beradi va nazariy asos va amaliy vosita bo'lib xizmat qiladi:

- *Chastotali tahlil*: korpus lingvistikada so'z, ibora yoki til birliklarining matnda yoki matnlar majmuasida (ya'ni korpusda) **qancha marta takrorlanganini aniqlash** jarayonidir. Bu tahlil orqali til birliklarining **foydalanish darajasi, ahamiyati, grammatik yoki leksik roli** aniqlanadi. So'zning har xil ma'nodagi qo'llanish chastotasini aniqlagan holda omonimlar tarjimasida to'g'ri ekvivalentni taqdim qiladi.
- *Kontekstual o'rganish*: So'zning yon-atrofidagi birliklar bilan birgalikda qanday semantik maydonda qo'llanilishi ko'riladi. Shu orqali omonim so'zning kontekstda qanday so'zlar yoki iboralar bilan kelganligiga qarab muqobil variant tanlanadi.
- *Tagging va lemmatizatsiya*: **Tagging** — bu matndagi har bir so'zga uning grammatik rolini (so'z turkumi, zamon, jins, son va b.) belgilovchi **teg** (tag) qo'shish jarayonidir. Omonimlar ko'pincha **bir xil shaklda, lekin har xil so'z turkumida** bo'ladi. Tagging yordamida bu **farqlar aniqlanadi**. Bu grammatik teglash **kontekstual tarjimada** to'g'ri variantni tanlashga yordam beradi. Omonimlar **turli grammatik shakllarda** bo'lishi mumkin, lekin ularning **asl shakli (lemma)** bir xil yoki farq qiluvchi bo'ladi. Lemmatizatsiya bu farqni aniqlashga yordam beradi.

Jahon tajribasi va korpuslar bilan ishlash. COCA (AQSh), BNC (Buyuk Britaniya), DeReKo (Germaniya), RNC (Rossiya), LCMC (Xitoy) kabi korpuslar omonimlarni aniqlashda keng foydalaniladi.

- COCA: zamonaviy ingliz tilidagi 560+ million so'z asosida omonimlarning kontekstual tahlili.
- BNC: 100 million so'zli Britaniya korpusi.
- DeReKo: nemis tilidagi omonimlarning sotsiokultural kontekstlarini tahlil qiladi.
- RNC: rus tilidagi "klyuch" so'zining turli ma'nolarini farqlashga yordam beradi.
- LCMC: mashinaviy o'rganish asosida xitoycha omonimlar kontekstini o'rganadi [Sharoff, 2006].

Sun'iy intellekt va WSD texnologiyalari orqali ham omonimlarning tarjimasining aniqliligini ta'minlash mumkin. WSD (Word Sense Disambiguation) – so'z ma'nosini kontekst asosida aniqlash texnologiyasi bo'lib, quyidagi yondashuvlar orqali amalga oshiriladi:

- Supervised Learning: Annotatsiyalangan ma'lumotlar asosida o'rganish.
- Pretrained Language Models: BERT, GPT, ELMo kabi neyron tarmoqlar asosida kontekstual semantik tafovutlarni aniqlash [Navigli, 2009].

Omonimlar tarjimasida ishtirok etadigan yana bir usullardan biri bu: lug'at bazalari va leksikografik tizimlar. Shular qatoriga quyidagilarni kiritish mumkin:

- WordNet [Fellbaum, 1998]: So'zlarni sinonimik to'plam (synset) asosida ajratadi.
- FrameNet: Semantik ramkalar orqali ma'no tafovutini ko'rsatadi.
- Open Multilingual WordNet: Ko'p tilli omonimlarni taqqoslash imkonini beradi.

Omonimlarning korpusga asoslangan lug'at bazalarini yaratish bir qator amaliy foydalanish imkoniyatlarini yaratib beradi. Quyidagi jadval sohalar va ulardagi istiqbolli imkoniyatlarni ko'rsatib beradi:

Soha	Foydasi
Mashinaviy tarjima	Omonimlarning to'g'ri tarjimasini tanlash
Chatbotlar / AI yordamchilari	Kontekstga mos javob berish
Til o'rgatish platformalari	Omonimlarni o'rgatishda interaktiv yondashuv
Ovozli interfeyslar	Noaniq so'z shakllarini aniq tushunish

Jahon tilshunosligida leksik-semantik jihatdan murakkab bo'lgan omonimlarning tarjimasini bir qator ilmiy va texnologik istiqbollarga yo'l ochadi. Bunga quyidagilarni misol qilish mumkin:

- Ko'p tilli korpuslar yaratish.
- Multilingual WSD tizimlarini ishlab chiqish.
- O'quvchilarga mo'ljallangan multimedia lug'atlar ishlab chiqish.

Texnologik jihatdan qaralganda, bu muammolarni hal qilish intellektual tarjima tizimlarining sifatini keskin oshirishga olib keladi. Omonimlarni to'g'ri tarjima qilish uchun zamonaviy mashina tarjimasini modellariga (masalan, Google Translate, DeepL) kontekstual tahlil imkoniyatlarini chuqurroq integratsiyalash talab etiladi. Shuningdek, bu soha ma'no ajratish (Word Sense Disambiguation — WSD) bo'yicha ishlanayotgan chuqur o'rganish (deep learning) modellari — ayniqsa BERT, RoBERTa, XLM-RoBERTa kabi kontekstual til modellarining rivojiga bevosita turtki beradi. Bu modellarda til birliklarining semantik aniqligi yuqori darajada ta'minlanadi, natijada omonimlarning tarjimasini ham aniqroq bo'ladi. Korpus vositalari, masalan, Sketch Engine, AntConc va Stanza kabi vositalar ham semantik kontekstga asoslangan avtomatik tahlil funksiyalar bilan boyitiladi. Bu esa leksikografiya, avtomatik tarjima, til o'rgatish dasturlarida yangi texnologik yechimlarni yuzaga chiqaradi. Bundan tashqari, omonimlar tarjimasini bilan bog'liq muammolarni hal qilish orqali ko'p tilli leksik resurslar, ontologiyalar va semantik ma'lumot bazalari (WordNet, BabelNet, Wikidata) kengaytiriladi, bu esa NLP tizimlarining yanada aqlli bo'lishiga imkon beradi.

Xulosa o'rinda shuni keltirish mumkinki, korpusga asoslangan yondashuvlar yordamida omonimlarni aniqlash, farqlash va tizimlashtirish hozirgi global tilshunoslikda dolzarb va istiqbolli yo'nalishdir. Korpus lingvistikasi asosida omonimlarning tarjimasini masalalarini hal etish nafaqat amaliy tarjima muammolarini bartaraf etadi, balki tilshunoslikning nazariy va amaliy sohalariga, leksikografiya hamda sun'iy intellekt va NLP texnologiyalariga innovatsion yondashuvlarni olib kiradi. Bu esa XXI asr tilshunosligining eng dolzarb yo'nalishlaridan biriga aylanadi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. O'zbek tili elektron korpusining kompyuter modellari (monografiya)./ Toshkent: Muharrir, 2021
2. Mirsaid, A., Bibigul, R., Altinbay, S., & Nilufar, A. (2018). Ontology of grammar rules as example of Noun the Uzbek language and Kazakh languages VI international scientific conference" Modern problems of the applied mathematics and information technology. Al Khorezmiiy-2018.
3. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
4. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In Компьютерная обработка тюркских языков. TURKLANG 2022 (pp. 231-240).
5. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
6. Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
7. Kilgarriff, A., & Grefenstette, G. (2003). *Introduction to the Special Issue on the Web as Corpus*. Computational Linguistics, 29(3), 333–347.

8. Navigli, R. (2009). *Word Sense Disambiguation: A Survey*. ACM Computing Surveys, 41(2), 1–69.
9. N. Z. Abduraxmonova. “Kompyuter lingvistikasi” Darslik. Toshkent – 2021
10. Sharoff, S. (2006). *Creating general-purpose corpora using automated search engine queries*. In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC).

ПРОБЛЕМА РАЗМЕТОК ПРИ ГЕНЕРАЦИИ КОРПУСА АБАЯ

Сабилова Д.А.
КазНПУ им Абая
Алматы, Казахстан

Аннотация: В статье дается обзор исследований последнего десятилетия, связанных с разработкой корпусов национальных языков, в том числе – поэтических субкорпусов. На сегодняшний день, как отмечают авторы, эта область корпусной лингвистики разработана недостаточно. В частности, освещены принципы отбора текстов для поэтических корпусов, типы меток в национальных корпусах языков и др. моменты, ключевые для предварительного этапа любого исследования, связанного с созданием и изучением национальных корпусов языков.

К трудностям, с которыми обычно сталкиваются разработчик курсов, авторами отнесены такие явления литературы, как скрытое авторство, коллективное, анонимное, народное авторство, а в отношении именно поэзии – частые нарушения метра, ритма, размера, рифмовки и строфики, что затрудняет машинную обработку поэтических текстов.

Авторы являются разработчиками поэтического корпуса Абая, входящего в казахский национальный языковой корпус. Отсюда интерес к поэзии Абая и к предшественникам, занимавшимся проблемами поэзии Абая, являющегося основателем не только литературной поэзии, но, прежде всего, казахского литературного языка.

Ключевые слова: национальные корпуса языка, разметка, морфологическая разметка, поэзия, поэтический корпус, Абай Кунанбаев.

Введение. Проблемы создания языковых корпусов является одним из приоритетных в современной лингвистике. Это направление позволяет собрать и осуществить классификацию всех текстов конкретного языка, предоставить площадку для продолжения исследований, использования ресурсов корпуса в практических учебных целях. Мы рассмотрим вопрос корпусного исследования стихотворного контента, то есть создания субкорпуса.

Национальная поэзия всегда привлекает внимание не только литературоведов, но и исследователей языка. Язык художественной литературы всегда был и является на сегодняшний день индикатором исторических явлений в языках.

К поэтическому корпусу, как известно, относятся элементы стихотворных художественных текстов. Он должен отображать особый уровень организации текста в рамках стиха. Автоматическая разметка текстов позволяет выделить различные аспекты изучения текстов на основе большого числа токенов. В рамках настоящего проекта мы исследуем морфологическую и поэтическую разметку. Рассмотрим, какие виды разметок, в каком направлении и каким образом осуществляются в исследованиях данного контекста. Классификация текстов осуществляется с жанровых позиций: это могут быть поэтические произведения, народные песни, лирические и эпические стихи, сонеты, хокку и многое другое. Каждому из этих жанров может соответствовать своя специфическая разметка, учитывающая особенности ритма, рифмы, метра, стиливых приёмов и других характеристик.

Морфологическая разметка в поэтическом корпусе позволяет выделить и классифицировать морфологические единицы, такие как слова, их формы, части речи, склонение, спряжение и другие морфологические признаки. Эта информация может быть полезной для анализа стихотворного текста и его особенностей. Поэтическая разметка, с другой стороны, помогает выявить поэтические приёмы, структуру стихотворения, его ритмические особенности, использование рифмы и другие поэтические элементы.

Например, в рамках разметки можно выделить долготу и краткость слогов, ударение, рифмованные схемы и т.д.

Исследования в области разметки поэтического корпуса могут направляться на различные цели. Некоторые исследователи интересуются анализом поэтической структуры и выявлением закономерностей в построении стихотворных текстов. Другие изучают стилистические особенности разных жанров. Также возможно проведение сравнительного анализа разных поэтических традиций или эпох. Разметка текстов может осуществляться различными способами, включая ручную разметку, использование специальных программ и алгоритмов, автоматическое распознавание особенностей текста и другие подходы. Важно подобрать наиболее эффективные методы разметки, которые соответствуют конкретным исследовательским вопросам и задачам.

Разметка поэтического корпуса может иметь несколько конкретных целей, в зависимости от направления исследования. Некоторые из них включают:

1. Анализ структуры и формы поэтических произведений: Разметка поэтического корпуса позволяет исследовать различные аспекты структуры и формы стихотворных текстов, такие как акцентуация, стихотворные размеры, рифма и др. Это может помочь выявить устоявшиеся паттерны и особенности в построении стихов, а также исследовать различные структурные вариации между разными жанрами поэзии или периодами литературы.

2. Анализ языковых особенностей и стилистики: Разметка поэтических текстов позволяет изучать различные языковые особенности, такие как использование образных средств, фигур речи, метафор, синтаксических конструкций и т.д. Это может быть полезно для анализа стилистики отдельных поэтов или возможности выявления общих черт в языковых особенностях той или иной эпохи.

3. Исследование эстетических аспектов поэтических произведений: Разметка поэтического корпуса позволяет анализировать и изучать эстетические элементы, такие как звучание, ритм, метрическая организация и т.д. Это важно для понимания того, как поэтический язык и структура создают определенное эстетическое впечатление на слушателя или читателя.

4. Создание инструментов для автоматического анализа и обработки стихотворных текстов: Разметка поэтического корпуса может быть использована для разработки и улучшения алгоритмов и программных инструментов для автоматического анализа и обработки стихотворных текстов. Это важно для создания инструментов, которые могут обрабатывать и классифицировать большие объемы стихотворных данных более эффективно и точно.

Исследования в области разметки поэтического корпуса имеют большое значение для различных дисциплин, таких как литературоведение, лингвистика, компьютерная лингвистика и др. Они помогают понять и анализировать различные аспекты поэтической речи и способы ее организации, что в свою очередь способствует более глубокому пониманию и интерпретации поэзии. В целом, исследования поэтического корпуса и его разметки позволяют более глубоко понять и анализировать поэтическую природу текстов, раскрыть их эстетические и стилистические особенности, а также провести широкий компаративный анализ разных поэтических традиций и жанров.

Методы.

Существует несколько методов исследования поэтического корпуса, которые могут быть применены для изучения различных аспектов литературной поэзии. Вот некоторые из них:

1. Качественный анализ: Этот метод включает подробное чтение и интерпретацию поэтических текстов с акцентом на качественное понимание их содержания, стиля и метафорического значения. Качественный анализ помогает выявить тематические и

стилистические черты, анализировать языковую особенность и исследовать организацию стихотворных форм.

2. Количественный анализ: Этот метод основан на статистическом исследовании текстов, чтобы найти закономерности и паттерны в поэзии. Количественный анализ может включать исследование частотности определенных слов, построение ворд-клаудов, анализ структуры и формы стихотворных произведений, а также использование компьютерных методов для анализа и интерпретации данных.

3. Социолингвистический анализ: Этот подход связывает стихотворение с социальными и культурными контекстами. Исследователи обращают внимание на языковые изыски, связанные с определенной группой людей или периодом времени. Социолингвистический анализ позволяет изучать, как язык и социальные факторы взаимодействуют в поэтическом тексте.

4. Контекстуальный анализ: Этот метод включает изучение поэзии в контексте других литературных произведений, исторических событий, культурных и социальных тенденций. Контекстуальный анализ позволяет понять, как поэтические произведения отражают и взаимодействуют с широким контекстом своего времени и места.

5. Компьютерное моделирование: В последнее время стало популярным использование компьютерных методов, таких как машинное обучение и обработка естественного языка, для анализа и моделирования поэтических текстов. Это позволяет автоматизировать и усовершенствовать исследовательские процессы, а также выявлять специфические особенности и закономерности в поэтическом корпусе.

Это только некоторые из методов, которые могут быть использованы для исследования поэтического корпуса. Конечный выбор метода зависит от конкретных целей и задач исследования. Комбинация различных методов может быть наиболее эффективной для получения полного и глубокого понимания поэтических текстов и их контекстов.

В ходе разработки поэтического корпуса нами используются различные методы в зависимости от конкретно решаемой задачи. Поскольку исследование находится на стыке лингвистики и литературоведения, при разработке поэтического корпуса Абая используются и следующие методы литературоведения: компьютерные методы исследования стиха, а также классические:

1. Метрический анализ: Изучение ритмической структуры стиха, включая количество слогов, ритмические схемы, ударения и так далее.

2. Рифмованный анализ: Анализ расположения рифм в стихотворении, выявление их типов (смежные, перекрестные и т. д.), роли в структуре текста.

3. Синтаксический анализ: Изучение синтаксических конструкций, порядка слов, предложений и их влияния на ритм и смысл.

4. Лексический анализ: Анализ использования слов, фраз и метафор в стихотворении для выявления образности, смысла и эмоциональной окраски.

5. Семиотический анализ: Исследование символов, знаков и метафорических образов в стихотворении и их значений.

6. Исторический контекст: Учет исторических обстоятельств, культурных и литературных тенденций времени написания стихотворения.

7. Компаративный анализ: Сравнение стихотворения с другими произведениями, а также схожих структур и тематик.

8. Постструктуралистский анализ: Анализ текста с учетом влияния социокультурных, политических и философских факторов на его интерпретацию.

Материал.

Последние десятилетия отмечены разнообразием исследований в области корпусов: Е.А.Гришина и др. [1] исследуют поэтический корпус (далее – ПК) в НКРЯ, Б.В.Орехов и Д.С.Степина [2] – поэтический корпус персидского языка, поэтические корпуса чешского (Колар), башкирского (Орехов). Если говорить о методах, то корпусный признается на

сегодняшний день оптимальным для изучения лингвистикой художественных текстов [66]. В дальнем зарубежье прецедент составляют Gutenberg English Poetry Corpus, gutenberg-poetry-corpus, Parsed Corpus of Old, English Poetry, The York-Helsinki Parsed Corpus of Old English Poetry и др.

Поэтический корпус как отдельный ресурс нами не найден. Он входит в состав Национального корпуса русского языка [3]. По адресу: <https://ruscorpora.ru/page/corpus-poetic-authors/> можно найти список авторов, чьи тексты обработаны в корпусе. Здесь указаны их имена, фамилии, даты рождения и смерти. В поэтическом корпусе насчитывается 3 млн слов. Поисковая система включает в себя лемму, словоформу, грамматические признаки, основную и дополнительную семантику (<https://ruscorpora.ru/search?search=CgQyAggJMAE%3D>). При введении искомого слова ресурс выдает все стихотворные тексты, содержащие данное слово. В разделах «конкорданс», «KWIC» и график соответственно представлены список употребления слова, ключевое слово в контексте, распределение результатов поиска по датам.

Е.А.Гришина и др. [1] отмечают некоторые особенности указания ссылок на авторство поэтических текстов. Они обусловлены тем, является ли автор единоличным творцом, является ли текст анонимным, представлен коллективом авторов, представлен ли автор (авторы) под псевдонимом или произведение народного авторства и нем может быть отмечено конкретным лицом. Также сложности вызывают переводы текстов. Также разработчиками русского поэтического корпуса указаны такие особенности создания ресурса, как обязательно указание первой строки стихотворного произведения, невозможность указания точных дат создания художественных текстов (в случаях, когда имеются ранние и поздние, окончательные редакции и другие сложные моменты).

Создатели корпуса также перечисляют все виды метрических форм, строф, на сегодняшний день принятых стиховедением, наличие клаузул, виды рифм, меры стиха, формула, включающая метр, число, меру и клаузулу, зоны рифмовки. Также в качестве прикладного момента исследователи отмечают методическую ценность корпуса, в частности, возможность решать лингвистические задачи на предмет истории ударения, русского произношения, лексики, а также стиховедческих задач, которые можно решать при помощи поэтического корпуса. Также, по мнению исследователей, корпус дает материал для исследования деграмматизации русской рифмы и много других возможностей использования корпуса. Таким образом, исследователи решают задачу, которую давно пытались решить ученые-исследователи русского стиха, а именно «создание некоторого предварительного варианта универсального электронного метрического справочника по всей русской поэзии – ресурса, о создании которого мечтало не одно поколение стиховедов и отсутствие которого по-прежнему существенно тормозит теоретические исследования русского стиха» [1, с.72].

К.М. Корчагин отмечает, что не все тексты стихов, написанных на русском языке, могут быть обработаны в корпус. Как правило, они проходят определённый отбор и могут быть разделены на три категории: наиболее значительные, оставившие определённый вклад в русскую поэзию и тексты автора, в единичном или немногочисленном виде представленные в корпусе. Автор относит к меткам поэтического текста комментарии двух видов: раскрывающие историю создания и структурно содержательные. Описание формальной структуры стиха включает следующие характеристики: «метр, длина строки, клаузула, строфика, (то есть окончание строки), строфика и ее графическое представление, способ рифмовки и разного рода дополнительные особенности структуры, учет которых может оказаться важным» [4, с.120]. Приведение таких характеристик предоставляет богатые возможности для исследований поэзии в историко-хронологическом или аспектном контекстах. В монографии «Корпусный анализ русского стиха» [5] Института русского языка имени В.Виноградова рассматривается семантика слов в стихах Батюшкова и других поэтов русской классики.

С.О.Савчук и А.А.Махова знакомят читателей с новым мультимедийным корпусом и обозначает принципы создания разметки. В мультимедийном корпусе нового поколения, разработанном авторами, включены не только вербальные тексты, но и аудио- и видеoverсии текстов. В частности, представлено прочтение поэтических текстов в трех вариантах: авторское, профессиональное и любительское. Весь контент разделен по родовой принадлежности и по типу исполнения. Также исследователи- составители корпуса выделяют следующие виды меток:

«1) разметка, общая для всех корпусов в составе НКРЯ (метатекстовая, морфологическая, семантическая аннотация); 2) разметка, используемая в поэтическом корпусе (метрические признаки, рифма), которая содержится в каноническом инварианте; 3) разметка, используемая в корпусах устной речи (акцентная, социологическая аннотация); 4) специфическая разметка просодического членения [Кривнова и др. 2016], которая наряду с акцентной и социологической аннотацией представлена в исполнительских версиях» [6, с.138].

Разработчики приводят пример сопоставления чтения стихотворения С.Есенина «Спит ковыль» Б.Галкиным и И.Ильиным, отмечая при этом такие варианты прочтения, как перестановка слов, различное интонационное оформление стиха, а также, с точки зрения авторов, наличие видео позволяет пользователям обратить внимание на жестовое сопровождение прочтения.

Д.В.Сичинава, обращаясь к особенностям работы с корпусами, отмечает, что это ведущий метод работы современной лингвистики и отмечает успехи в создании Британского национального корпуса, чешского корпуса Карлова университета и останавливается на проблемах создания корпуса ИППИ на основе Уппсальского корпуса. Ценным заявлением автора можно считать освещение подход к созданию корпусов, которые, по мнению автора, могут иметь следующий характер: полный (закрывающий в себе весь языковой контент национального языка), культурно-репрезентативный (содержащий культурно маркированный текст), «эталонный» (соответствующий узусу и языковой компетенции носителя). К основным сложностям автор относит «отсутствие единого стандарта подачи текста» [7], наличие опечаток, техническое несоответствие стандартам оформления текстов. Исследователь перечисляет этапы работы, состоящие из отбора текстов, перевод в единый формат, фильтрацию с целью минимизации омонимии, ручное снятие омонимии и приведение к единому Интернет- формату.

Морфологическая разметка, являющаяся центральным инструментом настоящего исследования, определяет частеречную принадлежность, начальную форму, в редких случаях – характеристика морфологической формы, встречающейся в тексте. Ориентировочную роль при этом играют теги – метки на ключевом слове. Грамматические теги при изучении корпусов призваны классифицировать и упорядочить корпус, объединить леммы. Процесс введения казахских текстов в национальный корпус будет последовательно осуществлён через введение предварительной разметки в текст в HTML-формате, обозначение морфологических признаков, метатекстовую разметку и конечной конвертации. Разметка – важный составляющий этап создания любого корпуса, которая означает информация, характеризующая отрезок текста в корпусе. Как известно, основное внимание создателей корпусов занимает синтаксическая разметка. В то время как морфологический уровень более понятен пользователю корпуса, не имеющему специального лингвистического образования.

Особенностью поэтической разметки является то, что она отображает особенности стиховой организации текста. В поэтическом корпусе русского языка исследована поэзия периода XVIII-XX веков. К.М.Корчагин [4] отмечает наличие двух разметок – грамматическую и стиховедческую. Такой подход к обработке текстов позволяет решать как лингвистические, так и литературоведческие задачи. Кроме того, в рамках литературоведения, как отмечает исследователь, корпус может содержать в себе интересы

как ученых, изучающих метрику, так и ученых, занимающихся историей поэзии и литературы, в целом.

Следующая прикладная проблема корпусной лингвистики в связи с созданием поэтических подкорпусов – отношение к поэтической иерархии, которая выделяет «образцовые» тексты и оттесняет незначительные или не столь значительные в социокультурном аспекте тексты. По мнению К.М.Корчагина, эта проблема требует осознанного подхода, поскольку для разных исследователей может быть интересна тот или иной поэтический контент.

Целью исследования, осуществляемого нами, является создание поэтического корпуса казахского языка посредством морфологической и поэтической разметки стихов Абая.

Проблема разметки чрезвычайно важна и актуальна на сегодняшний день. Для этого необходимо обратить внимание на особенности стихосложения Абая и казахского стихосложения в общем.

Казахское стихосложение на сегодняшний день является наименее исследованной областью теории казахской литературы. Первоначально эту категорию исследовал Ш. Уалиханов, а затем А. Байтурсынов дал ей научное обоснование в своей книге «Әдебиет танытқыш» (1926 г.). Байтурсынов выделил виды строф, размеры стихотворных строк, стихотворные течения и рифмы, а также предложил термины для различных аспектов стихосложения. Он также показал, что казахское стихосложение относится к силлабической стихотворной системе и выделил роль Абая как реформатора этой системы. Традицию развития стихосложения продолжили другие казахские поэты. В казахской поэзии широко используются различные стихотворные размеры, такие как семисложный, семи-восьмисложный и одиннадцатисложный.

Отличие ритмику слога и ритмики стиха А.Байтурсынов осуществил через термин «жорнақ». Им также было замечено, что он создается тремя способами в зависимости от природы языка. Первый способ применяется в языках с одинаковой кратностью гласных. Второй способ может быть применен в языках, где ударные слоги являются одинаковыми. И третий применяется в языках, в которых слоги не являются одинаковыми. А.Байтурсынов выделил следующие виды стиха:

1. «айшық» – виды стиха.
2. «бунақ» – промежуток между голосовыми волнами, который ощущается при чтении;
3. слог внутри стиха – «буын»;
4. промежуток между голосовыми волнами в стихе – «кезең»;
5. «тармақ» – стих;
6. «шумақ» – строфа;

А.Байтурсынов обосновал силлабическую систему казахского стихосложения. Также, отмечая роль Абая в развитии казахского стихосложения, он отметил его новаторскую роль в разработке размеров, строф и рифм, которые в дальнейшем были использованы казахскими поэтами.

Исследователь казахской литературы З.Ахметов отмечает явление чередования 7-ми и 8-мисложных строк в казахском стихосложении, в частности, в дастанах и толгау, где исполнение основано на речитативе – жедирме. Эпическая поэзия, по мнению исследователя, принимала форму жыр – больших поэтических форм. Кроме семисложников и восьмисложников в казахской поэзии также встречаются 11-ти сложные формы организации стиха. Горазде реже, по мнению исследователя, встречаются 6-ти и 4-сложные формы организации стиха.

Исследователи национального корпуса казахского языка отмечают ряд особенностей исследований в этой области. Так, А.К.Жубанов рассматривает назначение метатегов, которые делятся на внешние, структурные и технические. В состав тегов

включаются данные об авторе, выходные данные издания – источника текста, время обработки, исполнители, исходная кодировка электронной версии. Исследователь обращает внимание на актуальность стандартизации метатегов и отмечает существующие на сегодня: В настоящее время большое внимание уделяется вопросу стандартизации метатегов и отмечает существующие: TEI (Инициатива по кодированию текста), рекомендация EAGLES (Экспертная консультативная группа по стандартам языковой инженерии), Стандарт CES (стандарт кодирования корпуса), XCES (стандарт кодирования корпуса для XML), ISLE (Международные стандарты языковой инженерии), CDIF (формат обмена корпусными документами, BNC). А.К.Жубанов, анализируя опыт предшественников, отмечает успехи в создании корпус Брауна (BC) 1960 года, корпус, созданный в Уппсальском университете в Швеции, а также исследования Института языкознания им.А.Байтурсынулы в Казахстане.

Мы полностью согласны с исследователями, отмечающими, что «создание корпуса текстов является одним из основных требований, необходимых для полноценного развития государственного языка. Наряду с нормированием и стандартизацией языка он даёт представление о том, на каком уровне находился язык, словарный состав в определённый период, и об изменениях и новшествах в нём» [9].

Результаты и обсуждение.

Обоснованная нами исследовательская позиция находится в границах мировой корпусной лингвистики. Данная разработка не идет вразрез с работами по изучению поэтических корпусов различных языков, а путем генерации исследовательских идей авторов, перечисленных в исследовании [1-8]. В ходе работы проведен обзор имеющихся научных разработок по принципам составления языкового корпуса.

К предварительным результатам исследования можно отнести аспектный обзор исследований, затрагивающих проблему разработки корпусов национальных языков и поэтических субкорпусов в частности. В результате было выяснено, что поэтический корпус Абая – это электронный ресурс, содержащий все поэтические произведения Абая Кунанбаева, а также сопутствующие материалы, такие как варианты текстов (черновики, рукописи, публикации в различных изданиях с возможными комментариями издателей), научный корпус, словарь языка Абая. Вопрос о содержании корпуса остается дискуссионным и требует коррекции и обсуждения.

Недискуссионным же можно считать сам предмет исследования – творчество Абая, поскольку эта часть казахского культурного наследия может считаться уникальной как в содержательном, так и в воспитательном планах. Сохранение и продвижение стихов Абая играет ключевую роль в культурном обогащении и поддерживает связь между прошлым и настоящим. Строки Абая - это не только поэзия, но и уникальный лингвистический артефакт. Стихи поэта богаты уникальными оборотами, фразеологизмами и образами, отражающими дух и языковое наследие казахского народа. Сохранение этих стихов способствует сохранению языкового богатства и традиций народа для будущих поколений.

Стихи Абая являются не только произведениями искусства, но и историческими документами. Они отражают социальные, культурные и политические аспекты времени, в которое жил поэт. Сохранение этих стихов позволяет передать исторические знания будущим поколениям, обогащая их понимание собственной культурной истории. Поэзия Абая рассматривается как важная часть национальной идентичности казахского народа. В них заключены ценности, традиции и мудрость поколений. Сохранение этих стихов способствует укреплению самосознания и национальной гордости, служа основой для формирования культурной

Сохранение стихов Абая и исследование их в семантическом и других аспектах создает культурный мост между поколениями. Эти стихи являются неотъемлемой частью культурного диалога, способствуя пониманию и уважению традиций среди различных групп реципиентов культурной информации. Поэтическая разметка направлена на

выделение и аннотацию структурных элементов стихотворений. Это включает в себя выделение строф, строк, рифм и ритмических особенностей. Этот этап позволяет создать инструмент для анализа структуры и формы стихов Абая, а также для выделения характерных черт его поэтического стиля.

Результатом морфологической и поэтической разметки в поэтическом корпусе позволит ресурсу стать ценным источником для исследователей, лингвистов и литературных критиков. Этот корпус представит собой цифровой архив, удобный для хранения и доступа, обеспечивая беспрепятственное изучение стихов Абая

К результатам аспектного обзора можно отнести следующее:

1. Разработка поэтического корпуса имеет свои особенности, в отличие от основного корпуса, в частности, это связано с указанием на авторство, которое может быть сомнительным, коллективным или отсутствовать в принципе (например, в фольклоре), с определением точного метра, с переводными текстами и т.д.

2. На сегодняшний день имеется ценный исследовательский опыт по созданию национальных корпусов в английском, русском, чешском, персидском и других языках, в том числе – поэтических субкорпусов.

3. Создание поэтического корпуса Абая может быть отдельным ресурсом или входить в корпус казахского языка как его составляющая, как это представлено в русском национальном языковом корпусе. Однако такое нахождение зачастую ограничивает некоторые функции навигации по поэтическому корпусу, обусловленные спецификой стихотворной речи.

4. На сегодняшний день в казахском стиховедении нет достаточного количества работ по изучению особенностей стиха Абая, что создает дополнительные задачи для исследователей по уточнению некоторых стиховедческих категорий в поэзии Абая.

5. Разметка стихов Абая может исключить трудности, указанные выше при условии ограничения только морфологическим направлением при разработке меток.

Заключение

Поэтические корпуса являются важным инструментом в современной лингвистике, позволяя исследовать и классифицировать тексты конкретного языка. Создание субкорпусов поэзии открывает новые пути для изучения структуры и стилистики стихов. Морфологическая и поэтическая разметка позволяют выявить особенности текстов, такие как ритм, рифма, метр и другие поэтические элементы. Разметка текстов может быть направлена на анализ структуры и формы поэтических произведений, языковых особенностей и стилистики, а также эстетических аспектов.

Исследования поэтического корпуса имеют важное значение, помогая лучше понять и анализировать поэтическую природу текстов, раскрывая их эстетические и стилистические особенности. Различные методы исследования, такие как качественный анализ, количественный анализ, социолингвистический анализ, контекстуальный анализ и компьютерное моделирование, способствуют глубокому анализу стихотворных текстов и расширяют наше понимание поэзии.

Создание поэтических корпусов и их разметка требуют комплексного подхода, включая стандартизацию метатегов и выбор оптимальных методов разметки. Исследования в области поэтической разметки не только способствуют развитию лингвистики, но и открывают новые перспективы для изучения языковой культуры и литературного наследия.

Поэтический корпус также может содержать поэтическую разметку, которая отражает структуру стихотворения, особенности ритма, рифмы, метра и другие поэтические приемы. Например, такая разметка может включать информацию о стихической организации текста, наличие и расположение рифм, ударных и безударных слогов, использование фигур речи и другие характеристики, специфичные для поэтического стиля. Эти данные позволяют исследователям анализировать и сравнивать

стихотворные тексты, выделять сходства и различия между ними, изучать развитие литературных жанров и стилей в разные эпохи.

Одной из задач корпусного исследования стихотворного контента является автоматическое извлечение и разметка характеристик текста. Это может производиться с использованием компьютерных алгоритмов и программного обеспечения, которые позволяют автоматически обрабатывать большие объемы текстов и идентифицировать в них морфологические и поэтические особенности. Этот подход позволяет значительно ускорить и облегчить процесс исследования поэтического корпуса и обеспечить доступ к его ресурсам для различных целей, включая лингвистические и литературоведческие исследования, обучение в области русского языка и литературы, разработку компьютерных программ и искусственного интеллекта, создание новых художественных текстов и многое другое.

Таким образом, создание и исследование поэтического корпуса с морфологической и поэтической разметкой позволяет расширить наши знания о структуре и особенностях поэтического текста, проводить более глубокий анализ и сравнение стихотворных произведений, использовать эти данные в практических целях, таких как обучение и разработка программного обеспечения. Это важное направление в современной лингвистике и литературоведении, которое имеет множество перспективных исследовательских возможностей. Возможный поиск может быть осуществлен по многим параметрам, например, по конкретной словоформе, лексеме, морфологическим параметрам, линейной позиции относительно другой лексем/в предложении, дате и типу текста и т.п.

К проблемам разработки поэтических корпусов можно отнести моменты, как связанные с особенностями поэзии, так и не относящимися непосредственно к лингвистике или литературоведению. К ним можно отнести неоднозначность художественного слова, множественность интерпретаций, что затрудняет категоризацию текстов. Кроме того, поэтический текст трудно поддается автоматической обработке ввиду возможного смещения размеров, рифмы и других характеристик стиха. Художественно-образительные средства поэзии так же трудно подвергнуть автоматизации, а также они могут быть защищены авторскими правами и представлены не в таком большом количестве, как тексты других стилей и родов. Необходимо отметить и то, что потребителями. Для решения указанных проблем необходимы разработка методов автоматизированной обработки поэтических текстов с учетом их специфики, создание поэтических корпусов с тщательным аннотированием и категоризацией, расширение пользовательской аудитории поэтических корпусов.

Целью разработки казахского поэтического корпуса является создание доступного исследователям инструмента для глубокого анализа творчества Абая. Перспективы включают в себя расширение корпуса, интеграцию современных методов исследования, а также создание образовательных ресурсов для широкой аудитории. Этот проект направлен на укрепление языковой и культурной связи поколений, а также на вдохновение новых поколений на изучение творчества великого поэта.

Исследование финансируется в рамках проекта ИРН: BR21882334 «Разработка казахского поэтического корпуса: морфологическая и поэтическая разметка стихов Абая»

Использованная литература:

1. Гришина Е.А., Корчагин К.М., Плунгян В.А., Сичинава Д.В. Поэтический корпус в рамках Национального корпуса русского языка: общая структура и перспективы использования. [электрон ресурс].

<https://ruscorpora.ru/media/uploads/2022/04/21/k7.pdf>

2. Орехов Б., Степина Д. Персидский поэтический корпус.//Известия Института русского языка имени В.В.Виноградова, 2022.
3. Национальный корпус русского языка. электрон ресурс]. //http:// /ruscorpora.ru/
4. Корчагин К.М. Зачем нужен поэтический корпус и как его использовать. – Русская речь. – 2019. - №6. – С. 113-127.
5. Корпусный анализ русского стиха: Сборник научных статей Отв. ред. В. А. Плунгян, Л. Л. Шестакова. - М.: Издательский центр «Азбуковник», 2013.- 266 с.
6. Савчук С., Махова А. Мультимедийный поэтический корпус: принципы создания и особенности разметки. Известия Института русского языка имени В.В. Виноградова (2022)
7. Сичинава Д.В. К задаче создания корпусов русского языка. [электрон ресурс]. <http://www.mccme.ru/ling/mitrius/article.html>
8. Жұбанов А.Қ. Қазақ тілінің ұлттық корпусындағы іздестіру аппаратының ең құнды бөлігі – мәтіндерге тіркеліп берілетін метабелгіленім//Тілтаным. – 2016. – №2 – С.3-9.
9. Жанабекова А.А., Пирманова К.К., Токмырзаев Д.О. Рабочие механизмы «Корпуса текстов казахского языка» [электрон ресурс].
https://elibrary.ru/download/elibrary_48111249_34322030.pdf

LINGUISTIC DATABASE AND SOFTWARE OF MACHINE TRANSLATION

Mrs. S. Jothi

Assistant professor, Department of English,
Nadar Saraswathi College of Engineering and Technology
Email: jothirose32@gmail.com

Abstract: This paper presents the design and development of a linguistic database and accompanying software framework for machine translation (MT). The system integrates structured lexical, morphological, syntactic, and semantic data to support accurate and context-aware translation between multiple languages. The linguistic database is built using scalable, language-independent schemas that facilitate the inclusion of diverse language pairs and support rule-based, statistical, and neural MT approaches. The software suite provides tools for parsing, alignment, disambiguation, and synthesis, enabling high-quality translations through modular and extensible components. Emphasis is placed on interoperability, resource sharing, and customization, making the system adaptable for academic research, commercial applications, and low-resource language development. Initial evaluations demonstrate improved translation quality and reduced ambiguity in complex linguistic constructions, highlighting the effectiveness of the integrated linguistic resource.

This article explores the development of a comprehensive linguistic database and associated software designed to support machine translation (MT). The goal is to provide a scalable, modular, and language-independent framework that enhances the accuracy and contextuality of automated translations. The integration of morphological, syntactic, semantic, and lexical data into a unified system enables improved translation output, especially for complex and low-resource languages.

Keywords: Machine Translation, Linguistic Database, Rule-Based MT, Neural Machine Translation, Translation Software, Natural Language Processing, Computational Linguistics, Low-Resource Languages, Lexicon, Syntax.

Introduction

Machine translation has become a vital tool in bridging communication gaps across languages. However, the quality of translation depends heavily on the linguistic resources and computational models involved. While neural machine translation (NMT) systems have advanced significantly, the inclusion of structured linguistic data can further refine output quality. This paper presents the architecture and implementation of a linguistic database and software system aimed at enhancing machine translation capabilities. In today's interconnected digital landscape, the ability to communicate across language barriers has become more vital than ever. Whether for global business, academic exchange, tourism, or diplomacy, the demand for fast, reliable, and scalable language translation solutions continues to grow. This need has driven significant advancements in the field of **Machine Translation (MT)**—the automated process of translating text or speech from one language to another using software.

This paper explores the dual role of linguistic databases and software architecture in MT systems. It examines how linguistic resources are structured and utilized, the software tools that interface with these databases, and the ways in which these elements contribute to the overall effectiveness and adaptability of MT solutions. In doing so, the paper also reflects on the shift from rule-based to data-driven methods, the challenges in building multilingual linguistic resources, and the promising future of hybrid systems that integrate both rule-based and neural approaches.

Development of a Linguistic Database and Software for Machine Translation

Linguistic database is constructed with language-independent data structures, enabling support for a wide range of languages. It serves as a structured repository of language-related

information: lexicons, morphological rules, syntactic patterns, semantic roles, and translation equivalents. These databases are essential for analyzing, interpreting, and generating language in a meaningful and grammatically coherent way. While traditional Rule-Based Machine Translation (RBMT) systems are built almost entirely on these databases, newer paradigms like Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) have shifted towards corpus-based learning. Despite this, linguistic databases continue to play an auxiliary role in enhancing translation quality, particularly in low-resource languages and domain-specific contexts such as legal or medical translation.

The software component of MT systems involves complex pipelines. Preprocessing modules like tokenizers, POS taggers, and sentence parsers prepare the input for translation. The core translation engine, depending on its architecture (RBMT, SMT, or NMT), processes the input with or without explicit linguistic rules. Post-processing components help refine the translation by correcting grammar, adjusting formatting, and aligning the output to human-like fluency.

The core components include:

- **Lexical Database:** Contains entries with part-of-speech tags, lemmas, inflections, and usage frequencies.
- **Morphological Data:** Encodes rules for word formation and inflection.
- **Syntactic Structures:** Stores parsing rules and dependency relationships.
- **Semantic Mappings:** Links lexical items with ontological concepts and context-sensitive meanings.

1. Software Architecture

The software framework is built as a modular system with components for:

- **Preprocessing:** Tokenization, normalization, and sentence segmentation.
- **Analysis:** Morphological and syntactic parsing, part-of-speech tagging.
- **Translation Engine:** Hybrid model supporting rule-based, statistical, and neural approaches.
- **Post processing:** Reordering, disambiguation, and target language generation.

Conclusion

The development of a linguistic database and software for MT shows promising results in enhancing translation quality. Future work includes expanding language coverage, incorporating user feedback loops, and integrating with large-scale multilingual corpora and language models. Linguistic databases and MT software are two sides of the same coin. While neural networks have reshaped the MT landscape, the foundational role of linguistic knowledge remains essential. By combining robust linguistic data with cutting-edge software, the future of machine translation holds the promise of more accurate, inclusive, and context-aware communication.

This work underscores the continuous relevance of linguistic databases and smart software design in improving translation quality, enhancing cross-cultural communication, and ensuring inclusivity in the digital world. Though AI has transformed the MT landscape, the foundational linguistic infrastructure remains an indispensable part of any reliable translation system.

References:

1. Koehn, P. (2010). *Statistical Machine Translation*. Cambridge University Press.
2. Forcada, M. L., et al. (2011). "Apertium: a free/open-source platform for rule-based machine translation." *Machine Translation*.
3. Bahdanau, D., Cho, K., & Bengio, Y. (2014). "Neural machine translation by jointly learning to align and translate." *arXiv preprint*.
4. Tiedemann, J. (2020). "Massively Multilingual Neural Machine Translation." *Proceedings of COLING*.

Works Cited

1. Forcada, M. L., et al. (2011). "Apertium: A free/open-source platform for rule-based machine translation." *Machine Translation*, 25(2), 127–144. ► Describes the role of linguistic databases in RBMT systems like Apertium.
2. Koehn, P. (2010). *Statistical Machine Translation*. Cambridge University Press. ► A foundational text on SMT, explaining how linguistic resources are integrated into statistical models.
3. Bahdanau, D., Cho, K., & Bengio, Y. (2014). "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473. ► Introduces NMT and shows how neural models handle translation using aligned datasets.
4. Tiedemann, J. (2020). "Massively Multilingual Neural Machine Translation." *Proceedings of COLING 2020*. ► Discusses the use of multilingual databases in neural translation systems.

TIL TA'LIMIDA KOMPYUTER TEXNOLOGIYALARINING AHAMIYATI: CHATGPT, DUOLINGO VA IBRAT ACADEMY MISOLIDA

Hasanova Malika G'ofir qizi, Mirzo

Ulug'bek nomidagi O'zbekiston Milliy universiteti
Filologiya va tillarni o'qitish (o'zbek tili) yo'nalishi
2-bosqich talabasi

E-mail: hasanova_05@internet.ru

Annotatsiya: Ushbu maqolada raqamli texnologiyalarning til o'rganish jarayoniga ta'siri va ularning samaradorligi yoritilgan. Xususan, ChatGpt, Duolingo va Ibrat academy va boshqa ilovalarning til ta'limidagi ahamiyati xalqaro tadqiqotchilar fikrlari asosida tahlil qilinadi.

Kalit so'zlar: Til ta'limi, raqamli texnologiyalar, sun'iy intellekt, onlayn ta'lim, ChatGpt, Duolingo, Ibrat academ

Аннотация: В этой статье рассматривается влияние цифровых технологий на процесс изучения языка и их эффективность. В частности, важность ChatGpt, Duolingo, Ibrat Academy и других приложений в языковом образовании будет проанализирована на основе мнений международных исследователей.

Ключевые слова: языковое образование, цифровые технологии, искусственный интеллект, онлайн-обучение, chatgpt, Duolingo, Ibrat Academy

Abstract: This article examines the impact of digital technologies on the language learning process and their effectiveness. In particular, the importance of ChatGpt, Duolingo, Ibrat Academy and other applications in language education will be analyzed based on the opinions of international researchers.

Keywords: language education, digital technologies, artificial intelligence, online learning, chatgpt, Duolingo, Ibrat Academy

Kirish. Bugungi kunda raqamli texnologiyalar hayotimizni tubdan o'zgartirib borayotgani hammamizga ma'lum. Tabiiyki, ta'lim sohasidagi o'zgarishlar til ta'limining rivojlanishiga ham o'z ta'sirini o'tkazdi. Xususan, xorijiy tillarni o'rganish jarayonida kompyuter texnologiyalari yordamida ta'limni individuallashtirish, interaktivlashtirish va avtomatlashtirish imkoniyatlari kengaydi. Xalqaro tadqiqotlarga ko'ra, raqamli vositalar o'quvchining motivatsiyasini oshiradi va ta'lim samaradorligini sezilarli darajada yaxshilaydi ¹¹⁹. Ushbu maqolada ChatGpt, Duolingo va Ibrat Academy platformalarining til o'rganish va o'rgatishdagi imkoniyatlari ko'rib chiqiladi.

Asosiy qism. ChatGpt – sun'iy intellektga asoslangan chatbot bo'lib, OpenAI kompaniyasi tomonidan 2022-yilning noyabr oyida ommaga taqdim etilgan. ChatGpt OpenAI tomonidan tabiiy tilni qayta ishlash (NLP) sohasida ilg'or yutuqlarga erishish va foydalanuvchilarga inson tilida muloqot qilish imkonini beruvchi sun'iy intellekt tizimini yaratish maqsadida ishlab chiqilgan.¹²⁰ Godwin-Jones ta'kidlashicha, ChatGpt kabi vositalar talabalarning yozma savodxonligini oshirishda va grammatik xatolarni tahlil qilishda yangi imkoniyatlar yaratmoqda. U, shuningdek AI vositalari yordamida talaba o'ziga xos mashqlar yaratishi mumkinligini aytadi.¹²¹ Chatgpt sun'iy intellekt asosida ishlovchi platforma bo'lib til o'rganuvchilarga interaktiv suhbatlar, grammatik mashqlar, tarjima va yozma matnlar tahlili, kundalik va yillik topshiriqlar ro'yxati tuzish(deadline) va turli tilda tushunib suhbatlashish imkoniyatlarini taqdim etadi. Til o'rganuvchi uchun ChatGpt doimiy online, tezkor va cheksiz resursli virtual murabbiydir.

¹¹⁹ Blake, R. J. (2008). Brave New Digital Classroom: Technology and Foreign Language Learning. Georgetown University Press

¹²⁰ OpenAI(2022) **TechTarget. "What is ChatGPT?"**

¹²¹ Godwin-Jones, R. (2023). Artificial Intelligence in Language Learning. Language Learning & Technology Journal

Masalan, o'quvchi ingliz tilida essay yozadi va ChatGptdan feedback berishini so'raydi. Bu jarayon shaxsiy tahlil va xatolar ustida ishlashga ko'mak beradi.

Duolingo 2011-yilda Luis von Ahn va Severin Hacker tomonidan asos solingan. Platformaning asosiy maqsadi – ta'limni arzon, sifatli va butun dunyo bo'yicha hamma uchun ochiq qilish.¹²²

Duolingo- o'yinlar asosida tilni qiziqarliroq o'rganish imkonini beruvchi platforma bo'lib, foydalanuvchini har kuni 10-15 daqiqadan til o'rganishga undaydi. Bu ilova asosan bolalar va tilni endi o'rganayotgan o'quvchilar uchun bo'lib natijalarni ilovani o'zida statistik ko'rsatkichlar orqali kuzatib borish imkonini beradi. Ilmiy tadqiqotlarga ko'ra, Vesselinov va Grego tadqiqotida Duolingo ilovasida muntazam shug'ullanuvchi foydalanuvchilar to'rt haftada til ko'nikmalarining o'sishini namoyon qilgan. Ularning fikricha, o'yinlar o'rganishni kuchli motivatsiyalovchi omilga aylantiradi.¹²³ Bundan tashqari Duolingoda o'rganilgan mavzularni audio va matnli topshiriqlar orqali tekshirish va mustahkamlash mumkin. Statistika ma'lumotlariga ko'ra har kuni 15 daqiqa ilovadan foydalangan holda til o'rganish so'z boyligini 20% ga oshirish imkonini beradi.

Ibrat Academy – o'zbek tilidagi raqamli ta'lim platformasi bo'lib, o'zbeklar va o'zbek tilini o'rganayotgan chet ellik talabalar uchun zamonaviy va qulay til platformasi hisoblanadi. Ushbu milliy platforma 2022-yilda "Ibrat Farzandlari" loyihasi doirasida ishlab chiqilgan. Bundan ko'zlangan asosiy maqsad xorijiy tillarni o'rganishni ommalashtirish va ta'limni barcha uchun arzon, sifatli hamda ochiq qilishdir.¹²⁴ Ilovada nafaqat ingliz va rus tillari balki, 20 ga yaqin tillarni online video darslar, o'yinlar, mavzuni mustahkamlovchi topshiriqlar, testlar orqali o'rganish mumkin. Shaxsiy kuzatishlarim orqali, Ibrat Academy ilovasidan muntazam foydalanib 2 oy davomida ingliz tilidagi darajamni B1 dan B2 ga ko'tardim. Bundan ko'rinadiki, til ta'limidagi platformalar bosqichma-bosich o'z natijasini bermoqda.

Xulosa. Til ta'limidagi raqamli vositalarning rivojlanishi til o'rganishda yangi bosqichni boshlab berdi. ChatGpt – sun'iy intellektga asoslangan yondashuvni taklif etsa, Duolingo – gemmifikatsiya usuli orqali til o'ranishni osonlashtiradi, Ibrat Academy kabi milliy platformalar esa mamlakatimizda raqamli ta'limga bo'lgan talab yuqoriligi va samaradorli ekanini ko'rsatadi. Yuqorida sanab o'tilgan xalqaro tadqiqotlar shuni isbotladiki, raqamli vositalar til o'rganishda faqat yordamchi vositasini o'tamaydi balki, asosiy metodga ham aylanishi mumkin. O'zbekiston ta'lim tizimida ham bunday texnologik yondashuvlar kengroq qo'llanilsa, til o'rganuvchilarning til kompetensiyasi sezilarli darajada ortadi. Shu sababli kelajakda kompyuter lingvistikasi hamda sun'iy intellekt vositalari tatbiqi yo'nalishida ilmiy izlanishlar olib borish muhim hisoblanadi. Umumiy xulosa qilib aytganda, til ta'limida kompyuter texnologiyalaridan foydalanish kutilgan vaqtdan tezroq natija beradi. An'anaviy ta'limda kompyuter texnologiyalaridan keng va samarali foydalanish rivojlanishning muhim bosqichi sanalib, ilm-fan rivojiga katta hissa qo'shadi.

Foydalanilgan adabiyotlar ro'yxati:

1. Blake, R. J. (2008). Brave New Digital Classroom: Technology and Foreign Language Learning. Georgetown University Press.
2. OpenAI(2022) **TechTarget. "What is ChatGPT?"**
3. Godwin-Jones, R. (2023). Artificial Intelligence in Language Learning. Language Learning & Technology Journal
4. Von Ahn, L., Hacker, S. (2013). Duolingo: Learn a language for free while helping to translate the web. URL: <https://www.duolingo.com>
5. Vesselinov, R., & Grego, J. (2012). Duolingo Effectiveness Study. City University of New York.
6. Ibrat Farzandlari. (2023). Ibrat Academy haqida. URL:

¹²² Von Ahn, L., Hacker, S. (2013). Duolingo: Learn a language for free while helping to translate the web. URL: <https://www.duolingo.com>

¹²³ Vesselinov, R., & Grego, J. (2012). Duolingo Effectiveness Study. City University of New York.

¹²⁴ Ibrat Farzandlari. (2023). Ibrat Academy haqida. URL: https://uz.wikipedia.org/wiki/Ibrat_farzandlari

O'ZBEK TILINI IKKINCHI TIL SIFATIDA O'RGANISH JARAYONIDA TIL KORPUSINING ROLI VA SAMARADORLIGI

Normamatova Mashhura

O'zbekiston Milliy Universiteti 3-bosqich talabasi.

E-mail: mashhuranormamatova0@gmail.com

Ilmiy rahbar: **Abduraxmonova Nilufar**

Professori, filologiya fanlari doktori

Mirzo Ulug'bek nomidagi O'zbekiston milliy universiteti

E-mail: n.abduraxmonova@nuu.uz

Annotatsiya: Mazkur maqolada o'zbek tilini ikkinchi til sifatida o'rganishda korpus lingvistikasining o'rni va ahamiyati tahlil qilinadi. So'ngi yillarda ko'plab tillar, xususan ingliz tili uchun korpus asosida til o'rganish samarali metod sifatida tadqiq etilgan bo'lsa-da, o'zbek tilini ikkinchi til sifatida o'rganuvchilar uchun bu yondashuv asosida ishlab chiqilgan o'quv vositalari deyarli mavjud emas. Bu borada nazariy qarashlar bayon etilgan bo'lishiga qaramasdan, amaliy o'quv dasturlari ishlab chiqilmagan. Ushbu maqola korpus asosidagi yondashuvning nazariy asoslarini, boshqa tillarda qo'llanilgan tajribalarni va o'zbek tilida bu yondashuvdan foydalanish imkoniyatlarini tahlil qilishga qaratilgan. Shuningdek, mavjud o'zbek tili korpusining o'ziga xos jihatlari hamda ularning til o'rganishdagi potentsiali ko'rib chiqiladi.

Kalit so'zlar: korpusga asoslangan til o'rganish, til resurslari, tabiiy til materiallari, konkordans tahlili, so'z chastotasi, korpus asosidagi o'quvchi markazli yondashuv (data-driven learning).

Abstract: This article explores the role and significance of corpus linguistics in learning Uzbek as a second language. Although corpus-based language learning has been widely studied and successfully applied in teaching various languages—particularly English—there is still a noticeable lack of teaching materials developed on this basis for learners of Uzbek. Despite the presence of theoretical discussions, practical educational programs grounded in this approach are virtually nonexistent. The article aims to analyze the theoretical foundations of corpus-based approaches, review their application in other languages, and assess the potential of implementing this methodology in the context of the Uzbek language. It also considers the specific features of existing Uzbek language corpora and their applicability to language learning.

Keywords: corpus-based language learning, language resources, natural language data, concordance analysis, word frequency, data-driven learning.

Аннотация: В данной статье рассматриваются роль и значение корпусной лингвистики в процессе изучения узбекского языка как второго. Несмотря на то, что в последние годы обучение на основе корпусов активно исследуется и успешно применяется для изучения многих языков, в особенности английского, для изучающих узбекский язык соответствующих учебных материалов практически не разработано. Несмотря на наличие теоретических работ по данной тематике, практические образовательные программы, основанные на корпусном подходе, отсутствуют. Цель статьи — проанализировать теоретические основы корпусного подхода, рассмотреть опыт применения этого метода в других языках и оценить возможности его внедрения в преподавание узбекского языка. Также анализируются особенности существующих корпусов узбекского языка и их потенциал в обучении.

Ключевые слова: обучение на основе корпусов, языковые ресурсы, данные естественного языка, анализ конкордансов, частотность слов, обучение с опорой на данные.

So‘ngi yillarda ilm-fan, sport, iqtisodiyot kabi qator sohalarda erishilayotgan yutuqlar, bundan tashqari, O‘zbekistonda xorijiy talabalar sonining oshib borayotgani o‘zbek tilini chet tili sifatida o‘rganishga bo‘lgan ehtiyojni yanada oshirmoqda. Bu esa o‘zbek tilini ikkinchi til sifatida o‘rganuvchilar uchun mustaqil o‘rganishga yordam beruvchi samarali metodlarni ishlab chiqishni dolzarb masalaga aylantiradi. Davr talabi bilan bu jarayonda an‘anaviy usullar bilan bir qatorda zamonaviy texnologiyalarga asoslangan yondashuvlar, xususan, til korpuslaridan keng foydalanish til o‘rganish jarayonida yangi imkoniyatlar yaratadi.

Bugungi kunda o‘zbek tilini ikkinchi til sifatida o‘rganuvchilar uchun tizimli va zamonaviy metodik materiallar yetarli emas. Mavjud o‘quv qo‘llanmalarining asosini esa grammatik qoidalar, tarjima asosidagi mashqlar va sintaktik konstruksiyalarni o‘rgatishga qaratilgan birliklar tashkil etadi. Bunday yondashuv o‘rganuvchilarning kommunikativ kompetensiyasini rivojlantirishda, ayniqsa, real nutq sharoitida qo‘llaniladigan birliklarni o‘zlashtirishda yetarli darajada samara bermaydi.

Zamonaviy til o‘rgatish metodologiyasida esa kontekstual va funksional yondashuvlar, xususan, korpus lingvistikasi vositalaridan foydalanish muhim ahamiyat kasb etmoqda. Til korpuslari asosida tuzilgan o‘quv materiallari o‘rganilayotgan tilning real qo‘llanilishiga asoslanib, o‘rganuvchiga tabiiy til muhitiga yaqinlashishiga yordam beradi. Bu nafaqat lug‘aviy boylikni, balki kollokatsiyalar, uslubiy variantlar, nutq aktlaridagi turlicha qo‘llanishlarni ham o‘zlashtirish imkonini beradi.

Jahon tilshunosligida korpuslardan til o‘rganish va o‘qitish sifatida foydalanish 1980-yillardan boshlangan edi. Aynan shu davrda korpuslar zamonaviy ko‘rinishda – tizimli tashkil etilgan elektron matnlar to‘plami sifatida shakllandi. Kompyuter texnologiyalarining rivojlanishi bilan birgalikda korpus lingvistikasi til o‘qitish va tadqiqot sohasida o‘z o‘rnini egalladi. Barlow (1996:2) ta’kidlaganidek: “Korpuslar o‘quvchi o‘zlashtirishi kerak bo‘lgan til andozalarining nafaqat ko‘lamini, balki ularning chastotasini ham ko‘rsatadi. Bu esa o‘quv materiallari va dasturlarni ishlab chiqishda muhim omildir. Bundan tashqari, korpuslardan foydalanish ikkinchi tilni o‘rganish sohasi, balki butun tilshunoslikka tub o‘zgarishlar olib kirish salohiyatiga ega”. Ushbu fikrdan kelib chiqilsa, korpus asosida tuzilgan darsliklar hayotda ko‘p ishlatiladigan va real misollarga asoslanadi.

Bugungi kungacha jahon tilshunosligida korpusga asoslangan til o‘rganishning samarali usullarini ishlab chiqish uchun bir qancha tadqiqotlar o‘tkazilgan. Jumladan, Boontam va Phoocharoensil (2018) o‘zlarining tadqiqotini 4-sinf o‘quvchilaridan iborat 30 ta yosh tailandlik o‘quvchilar bilan olib borgan. Ular ingliz tilidagi 3 ta predlogni Data-driven Learning (DDL) usulida o‘rganish samaradorligini tadqiq etishni maqsad qilgan. Tadqiqotning yana bir maqsadi, o‘quvchilarning DDL faoliyatlari orqali o‘rganishdagi munosabatlarini tahlil qilish edi. Tadqiqot 6 hafta davom etgan va ishtirokchilar haftasiga bir marta 50 daqiqalik qo‘shimcha darsda qatnashganlar. Ishtirokchilar 2 xil tes topshirgan: pre-test va post-test. Har ikkala test bir xil darajadagi savollardan iborat bo‘lib, ularda 15 ta bo‘shliqni to‘ldirish, 15 ta Grammatik xatolikni aniqlash va 5 ta jumla tuzish topshiriqlari bo‘rdi. Pre-test natijalari bilan solishtirganda, post-test natijalari sezilarli yaxshilanishni ko‘rsatdi. Tadqiqotchilar bu natijani DDL faoliyatlarining 6 hafta davomidagi ta’siri deb hisoblashgan. Shuningdek, ishtirokchilar DDL usulida o‘rganishga ijobiy munosabatda bo‘lib, uni foydali va qiziqarli deb ta’riflashgan.

Çelik va Elkatmiş (2013) tomonidan o‘tkazilgan eksperimental tadqiqotda turkcha tinish belgilarini o‘rganishning ikki usuli—korpus asosidagi usul va ma’ruza asosidagi usul taqqoslangan. Tadqiqot til o‘rganuvchilarini korpusga asoslangan til o‘qitish usuliga ekspozitsiya qilishning tinish belgilari o‘rganishiga ta’sirini tadqiq etishni maqsad qilgan edi. Tadqiqotga jami 171 nafar universitet talabalari Kirikkale universitetining Boshlang‘ich ta’lim fakultetidan jalb qilindi. Tadqiqotda test qilinadigan ko‘nikmalar: nuqta, vergul, nuqtali vergul, ikki nuqta, uch nuqta, undov belgilari, qo‘sh tirnoq, apostrof va defies edi. Tadqiqot pre-test va post-test nazorat guruhlari dizaynida amalga oshirildi. Tadqiqotda ikki eksperimental va ikki nazorat guruhi qatnashdi. Ma’lumot to‘plash vositalari sifatida tinish belgilari bo‘yicha yutuq testi, korpusga

asoslangan til o'rganish yondashuvi bo'yicha so'rovnoma va yarim tuzilgan intervyu shakllari ishlatilgan (eksperimental guruhlardan 10 nafar ishtirokchi bilan intervyular o'tkazilgan). Post-test natijalari, korpusga asoslangan til o'qitish usulini o'rganayotgan eksperimental guruhlarining, ma'ruza asosidagi o'qitish metodida o'qigan nazorat guruhlari bilan solishtirganda, yaxshiroq natijalarga erishganligini ko'rsatdi. Tadqiqot natijalariga ko'ra, eksperimental guruhlardagi o'quvchilarning aksariyati bu usulni ijobiy baholagan va tinish belgilari bo'yicha korpusga asoslangan usulda o'rganishni foydali va samarali deb hisoblagan.

Levchenko (2017) Rossiyaning Samara universitetining 10 nafar doktorantlari (6 erkak, 4 ayol) orasida tadqiqot o'tkazdi. Tadqiqotning maqsadi ingliz tilini akademik maqsadlar uchun o'rgatishda (EAP) korpusga asoslangan sinf faoliyatlarini (C-BCA) qo'llash haqida akademik xabardorlikni oshirish edi. Tadqiqot yozish kompetensiyalarini rivojlantirishga qaratilgan edi. Metodik manba sifatida tadqiqotchi Academic Word List (AWL) ni ishlatdi, unda 570 ta so'z oilasi mavjud (Coxhead tomonidan 2000 yilda ishlab chiqilgan). Tadqiqot 15 hafta davom etdi. Ishtirokchilar haftada ikki marta, har biri ikki soatdan darslarda qatnashdilar. Ishtirokchilar tarix, sotsiologiya, psixologiya, pedagogika, boshqaruv va iqtisodiyot kabi turli sohalaridan bo'lgan. Ular ingliz tilidagi darslarga postgradua imtihonini topshirish va ilmiy sohalarida tadqiqot maqolalarini chop etish, taqdimotlar qilish va muloqot ko'nikmalarini yaxshilash uchun tayyorlanishayotgan edi. O'rganish jarayonida ba'zi talabalar onlayn ishlashdi, boshqalari esa korpusning bosma nusxalaridan foydalandi. Ishtirokchilarga tadqiqotning uchta bosqichida (boshida, o'rtada va yakunda) so'rovnoma berildi. Natijalarga ko'ra, ishtirokchilar korpusga asoslangan yondashuvni ingliz tilidagi akademik ko'nikmalarini rivojlantirish uchun juda foydali deb baholadilar. Tadqiqotning asosiy e'tibori akademik yozish ko'nikmalarini rivojlantirishga qaratilgan bo'lib, ishtirokchilar treningdan so'ng kelajakdagi yozma ishlarida qo'shishlari kerak bo'lgan yuqori chastotali akademik so'zlarni o'zlashtirdilar.

Ko'rib chiqilgan tadqiqotlar, korpusga asoslangan metod qanday qilib til o'rganish va o'qitish sohasiga, o'quvchilarning malaka darajasidan qat'iy nazar, hissa qo'shishi mumkinligini ko'rsatdi. O'qitish va o'rganish sohasida, korpus texnikalari o'quvchilar va o'qituvchilar uchun ayniqsa foydalidir. Korpus o'quvchilarning savollariga javob beradi va o'qituvchilarga o'qitish materiallarini qiziqarli va haqiqiy tarzda tayyorlash imkonini beradi. To'g'ri boshqarilganda, u moslashuvchan o'rganish strategiyalarini taqdim etadi. Ko'rib chiqilgan tadqiqotlar asosida, korpusga asoslangan metodning til o'rganish va o'qitishda samarali ekanligi aniqlandi. Xususan, DDLda, o'quvchilar ma'lumotlarga to'g'ridan-to'g'ri kirish imkoniyatiga ega bo'lib, bu juda samarali va tez o'rganish natijalariga olib keladi. Korpuslar til o'rganish va o'qitishda farq yaratishi mumkin, bunda o'quvchilar lingvistik tadqiqotchilarga, o'qituvchilar esa jarayonning yordamchilari, yo'riqchilari va koordinatori bo'lib qoladilar.

Biroq, tadqiqotlarning ko'pchiligi ingliz tilini o'rganishga qaratilgan bo'lib, korpus asosidagi o'rganishning boshqa tillarda, xususan, o'zbek tilida qanday natijalar berishi to'g'risida yetarlicha dalil mavjud emas. Bu, o'z navbatida, ushbu maqolada yoritilayotgan masalaning dolzarbligini oshiradi. O'zbek tilini ikkinchi til sifatida o'rganishda qanday qiyinchiliklar mavjudligi, mavjud darsliklar va o'qitish metodikasida qanday bo'shliqlar borligi va korpus texnologiyasi bu muammolarni qanday hal eta olishi bugungi kunda kengroq tadqiqot talab qiladigan mavzulardandir.

Shu bois, ushbu maqola aynan o'zbek tilini o'rganishda korpus asosidagi yondashuvning zarurligini asoslash, uning metodik afzalliklarini ochib berish hamda bu sohada olib boriladigan keyingi tadqiqotlar uchun nazariy asos yaratishni maqsad qiladi.

Foydalanilgan adabiyotlar:

1. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. *Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL)*, 6(1-2019), 131-7.

2. Abdurakhmonova, N., Tulyev, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: *uzbekcorpus. uz*. In *2021 International Conference on Information Science and Communications Technologies (ICISCT)* (pp. 1-4). IEEE.
3. Abdurakhmonova, N. (2016). The bases of automatic morphological analysis for machine translation. *Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta*, 2(38), 12-7.
4. Suleymanov, D., Nevzorova, O., Gatiatullin, A., Gilmullin, R., & Khakimov, B. (2013). National corpus of the Tatar language "Tugan Tel": grammatical annotation and implementation. *Procedia-Social and Behavioral Sciences*, 95, 68-74.
5. Abdurakhmonova, N., Tulyev, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: *uzbekcorpus. uz*. In *2021 International Conference on Information Science and Communications Technologies (ICISCT)* (pp. 1-4). IEEE.
6. Abdurakhmonova, N., Tulyev, U., & Gatiatullin, A. (2021). Linguistic functionality of Uzbek Electron Corpus: *Uzbekcorpus. uz*. In *International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities, ICISCT 2021*.
7. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). *Uzwordnet: A lexical-semantic database for the uzbek language*. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
8. Barlow, M. (1996). Corpora for theory and practice, *International Journal of Corpus Linguistics*, 1(1), 1-38.
9. Boontam, P and Phoocharoensil, S. (2018). Effectiveness of English preposition learning through Data-Driven Learning (DDL). *3L: The Southeast Asian Journal of English Language Studies – Vol 24(3)*, 125-141.
10. Çelik, S. and Elkatmiş, M. (2013). The Effect of corpus assisted language teaching on the learners' proper use of punctuation marks, *Educational Sciences: Theory & Practice*, 13(2), 1090-1094.
11. Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Addison Wesley Longman Inc.
12. Levchenko, V. (2017). Use of Corpus-based classroom activities in developing academic awareness in doctoral students, *The New Educational Review*, 48, 28-40. DOI: 10.15804/tner.2017.48.2.02
13. Vyatkina, N. and Boulto, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology*, 21(3), 1-8.

РАЗРАБОТКА КАЗАХСКОГО ПОЭТИЧЕСКОГО КОРПУСА: МОРФОЛОГИЧЕСКАЯ И ПОЭТИЧЕСКАЯ РАЗМЕТКА СТИХОВ АБАЯ

Мырзаханова Гулжан Куттыбаевна

Старший преподаватель, магистр,
Казахский национальный педагогический
университет имени Абая,
Казахстан, Алматы
ID проекта: IPH: BR21882334

Поэзия Абая Кунанбайулы отличается глубокой философской наполненностью и своеобразной ритмико-метрической структурой, которая играет ключевую роль в поэтическом и эстетическом звучании его произведений. Цель данного исследования — изучение ритмических структур поэзии Абая с помощью лингвистического моделирования, с акцентом на метрические схемы и их связь с семантическими и стилистическими элементами его стихов.

Исследование основано на корпусе поэзии Абая, который был оцифрован и аннотирован с учетом лингвистических и просодических характеристик. С использованием вычислительных инструментов, таких как анализаторы ритма, детекторы слоговой структуры и модели идентификации ударений, в исследовании анализируются метр и ритм произведений поэта. Рассматриваются различные метрические типы, такие как ямб, хорей и анапест, с целью выявления предпочитаемых ритмических структур в поэтическом стиле Абая.

Анализ показывает, что Абай использовал разнообразные ритмические схемы для усиления эмоционального и философского воздействия своих произведений. Ритм в его поэзии служит не только формальным элементом, но и средством усиления смысла, акцентирования ключевых идей и вызова определённых эмоциональных откликов. Кроме того, ритмическое разнообразие в поэзии Абая отражает его способность сочетать традиционную казахскую просодию с более современными поэтическими формами, что способствует уникальности его поэтического голоса.

Результаты исследования дают представление о взаимосвязи между лингвистическим ритмом и смыслом в произведениях Абая. Посредством компьютерного моделирования ритмических структур работа демонстрирует, как ритм функционирует как стилистическое средство в казахской поэзии и вносит вклад в общее эстетическое восприятие текста читателем.

Ключевые слова: поэзия Абая, ритмическая структура, лингвистическое моделирование, корпусная лингвистика, поэтический метр, цифровой корпус, морфологическая разметка.

ЯЗЫКОВОЙ РЕСУРС ПОЭТИЧЕСКИХ КОРПУСОВ – ЦИФРОВОЙ РЕСУРС ДЛЯ АНАЛИЗА ПОЭЗИИ

Мадиева Гульмира

доктор филологических наук, профессор

Саткенова Жайна

кандидат филологических наук, доцент

Алимтаева Ляззат

кандидат филологических наук

Казахский национальный университет им. аль-Фараби

Annotatsiya. Maqolada she'riy korpusning tilni saqlash, o'qitish va ilmiy tadqiqotlar uchun qimmatli raqamli resurslar sifatidagi imkoniyatlari, shuningdek, uni turli nuqtai nazardan o'rganish, shu jumladan she'riy nutqni tahlil qilishning funksional imkoniyatlari ko'rib chiqiladi.

Kalit sozlar. til, tabiiy til, Milliy til korpusi, poetik korpus, korpus lingvistikasi.

Abstract. The article examines the potential of poetic corpora as valuable digital resources for language preservation, teaching, and scientific research, as well as the functional capabilities of studying it from various points of view, including the analysis of poetic speech.

Keywords: language, natural language, National Language Corpus, poetic corpus, corpus linguistics.

Введение

На протяжении многих столетий учеными изучается такое явление, как язык. В общем смысле язык – «сложная знаковая система, естественно или искусственно созданная и соотносящая понятийное содержание и типовое звучание/написание. Термин «язык», понимаемый в широком смысле, может применяться к произвольным знаковым системам, хотя чаще он используется для более узких классов знаковых систем.» [1]. Естественный язык (вербальный, человеческий язык), исторически сложившийся естественным путем в процессе социальной, когнитивной деятельности людей, используется определенной этнической группой и может быть распространен на определенной территории.

Естественный язык – это своеобразная знаковая система, которая создана как инструмент общения, является средством идентификации каждого народа. По статистике 2025 г. в мире насчитывается более 7 159 языков, в которые входят и наречия, <...> «около 44% всех языков сейчас находятся под угрозой исчезновения, часто с менее чем 1000 оставшимися пользователями. Между тем, 20 крупнейших языков мира являются родными языками для более чем 3,7 миллиарда человек в общей сложности. Это всего лишь 0,3% языков мира, на которых говорит почти половина населения мира!» [2]. Количество языков является изменчивой переменной. В результате использования инструмента измерения EGIDS получены следующие данные: «Число живых языков сокращается со средней скоростью 1 язык в две недели. На 40 наиболее распространенных языках разговаривает примерно 2/3 населения Земли» [3]. Языки характеризуются многими параметрами: структурой, системностью, количеством носителей этого языка, распространенностью, функциональностью, витальностью и др. В разных странах принимаются соответствующие меры по сохранению языков.

Основная часть. Одними из таких инструментов сохранения и изучения являются Национальные корпуса языков, в том числе поэтические корпуса/подкорпуса, которые в современное время являются ценным результатом цифровизации культурного наследия

этнического многообразия. Представим несколько аспектов, которые определяют функциональные возможности национального корпуса:

1. Сохранение языкового наследия

Корпус фиксирует язык в его разных исторических и функциональных проявлениях – от древних текстов до современного сленга. Благодаря этому:

- сохраняются формы и конструкции, которые выходят из употребления;
- можно проследить эволюцию слов и грамматических норм;
- поэтический корпус фиксирует уникальные авторские идиостили и метафорические модели.

2. Исследование и описание языка

Научные исследования на основе корпуса помогают:

- описывать грамматику и лексику более точно и объективно;
- отслеживать живую речь и ее изменения;
- выявлять закономерности употребления слов и выражений, в том числе в поэзии;
- представлять статистические данные по использованию той или иной языковой формы, выражения в определенных ситуациях.

3. Образование и просвещение

Корпус используется в:

- преподавании языка, в том числе как иностранного;
- разработке учебников, пособий, справочников, словарей различных жанров, тезаурусов;
- повышении как устной, так и письменной языковой грамотности.

4. Поддержка литературного языка

Поэтический корпус особенно важен для сохранения культурного и эстетического аспекта языка:

- демонстрирует богатство выразительных средств;
- сохраняет уникальные поэтические тексты, часто плохо доступные в других форматах;
- стимулирует интерес к чтению и изучению поэзии.

Если рассматривать этот вопрос шире, можно говорить о том, что корпус – это своего рода "языковая архивация", дающая возможность как современным, так и будущим поколениям видеть язык в диахронии и синхронии, т.е. каким был язык в разные эпохи, как говорили поэты, писатели, журналисты, и как говорят в настоящее время.

Например, в корпусе можно исследовать какое-либо слово или поэтический образ наглядно и интересно. В качестве примера сравним, как русский и казахский языки сохраняются и исследуются с помощью корпусов, особенно с точки зрения поэтических текстов (см. Таблицу 1).

Таблица 1 Представленность русского и казахского языков в корпусах

Аспект	Русский язык (НКРЯ)	Казахский язык
Корпус	Национальный корпус русского языка (в т.ч. Поэтический корпус; https://ruscorpora.ru/)	Национальный корпус казахского языка (Ұлттық корпустар: https://qazcorpus.kz/ ; http://web-corpora.net/KazakhCorpus/ ; https://corpus.kazlang.kz/)
Объем	Миллионы текстов (литературных, научных, разговорных, поэтических)	Меньший объем, но активно развивается с 2012-го года
Поэзия	От древнерусской поэзии до современных авторов	Включает как классиков (Абая), так и современных поэтов

Аспект	Русский язык (НКРЯ)	Казахский язык
Интерфейс и функции	Удобный поиск по словам, метке времени, автору, жанру, грамматическим тегам	Аналогичный функционал, но пока менее продвинутый по удобству
Роль в сохранении языка	Позволяет отслеживать изменения, анализировать стили, сохранять редкие формы	Критически важен для сохранения казахского языка, особенно в условиях билингвизма

Примеры поэтического поиска

Пример 1. Слово "душа" в русской поэзии

В НКРЯ (<https://ruscorpora.ru>), в Поэтическом корпусе, можно найти десятки тысяч употреблений слова "душа":

*И скучно и грустно, и некому руку подать
 В минуту душевной невзгоды...*
 (М. Лермонтов)

Корпус покажет:

- частотность слова "душа";
- с какими словами оно чаще сочетается (например, "томится", "плачет", "поет");
- эволюцию употребления от XIX до XXI века.

Пример 2. Слово "жүрек" (сердце) в казахской поэзии

В Казахском корпусе (<https://corpus.kazlang.kz>), можно найти строки из поэзии Абая:

*Жүрек – теңіз, қызығу – шетсіз аспан,
 Түсіне алсаң, сырын айтпас, жай ашпас.*
 (Абай Құнанбайұлы)

Корпус позволяет, кроме вышеперечисленных:

- находить идиоматические выражения, связанные с «жүрек»;
- анализировать, как изменялась метафорика сердца в казахской поэзии.

Что это дает?

• **связь с культурой:** через корпус мы видим, как формировались ключевые образы – «душа» в русском языке и «жүрек» в казахском.

• **языковая грамотность:** сочетаемость этих слов с другими словами, умение использовать в определенном контексте, умение использовать эти понятия в эмоциональном, выразительном аспекте, особенно важно для преподавания языка как родного и как второго.

• **оцифровка идентичности:** особенно для казахского языка, где корпус – инструмент не просто анализа, но и сохранения подлинной культурной памяти.

Обобщенный анализ ключевых концептов на основе доступных данных и исследований дает следующие результаты:

Эволюция слова «душа» в русской поэзии

Слово «душа» является одним из ключевых концептов в русской поэзии, особенно в лирике. Его использование отражает культурные и философские изменения в обществе.

- В поэзии XIX века «душа» часто ассоциировалась с религиозными и духовными поисками. Такие поэты, как А.С. Пушкин и М.Ю. Лермонтов, использовали этот образ для выражения внутренних переживаний и моральных дилемм.

- В XX веке, в советскую эпоху, акцент сместился на социальные и коллективные аспекты. Тем не менее, поэты продолжали использовать образ «души» для выражения личных чувств и связи с природой.

- В XXI веке современные поэты часто переосмысливают понятие «душа», связывая его с индивидуальной идентичностью и психологическими аспектами.

Эволюция слова «жүрек» в казахской поэзии:

В казахской поэзии слово «жүрек» (сердце) также имеет глубокое символическое значение, отражающее эмоциональные и культурные ценности.

- В классических произведениях, таких как поэзия Абая, «жүрек» символизирует центр чувств и моральных качеств.

- Современные казахские поэты продолжают использовать образ «жүрек» для выражения личных эмоций, любви и национальной идентичности.

Оба слова – «душа» и «жүрек» – играют центральную роль в поэтических традициях русского и казахского языков, соответственно, отражая эволюцию культурных и философских представлений. Несмотря на то, что точные данные по частотности по десятилетиям в настоящее время недоступны, можно с уверенностью сказать, что эти концепты остаются актуальными и значимыми в литературе.

Это только один из примеров как можно использовать поэтический корпус в качестве цифрового ресурса.

Поэтические корпуса представляют собой специализированные электронные коллекции поэтических текстов, собранные и размеченные с целью автоматизированного анализа, хранения и исследования поэзии. Эти ресурсы играют важную роль в цифровой филологии, корпусной лингвистике, стилистике и литературоведении. В отличие от общетекстовых корпусов, поэтические корпуса обладают уникальными особенностями, связанными со спецификой поэтического языка, метрической структурой и рифмой.

Создание поэтических корпусов преследует несколько целей:

- анализ структуры стиха: метрика, рифма, строфика, синтаксис;
- стилистическое и лексико-грамматическое исследование: выявление частотных моделей, тропов, эпитетов;
- сравнительное литературоведение: сопоставление языка, стиля, художественного мировоззрения поэтов, направлений, эпох;
- обучение нейросетей и ИИ: генерация поэзии, автоматическое распознавание жанров и авторов;
- сохранение культурного наследия: цифровизация и доступ к редким поэтическим текстам.

В мировой практике существуют такие поэтические корпуса, как:

1. Project Gutenberg Poetry Corpus

Свободная коллекция англоязычных поэтических текстов из проекта Гутенберг. Используется для обучения моделей машинного обучения.

2. Corpus of Historical American English (COHA)

Несмотря на то, что это общий корпус, он содержит обширные поэтические тексты XIX–XX веков, полезные для изучения исторического развития поэтического языка.

3. Русский поэтический корпус (РПК)

Проект Института русского языка РАН содержит стихи русских поэтов XVIII–XX веков с разметкой рифмы, метра, частей речи и синтаксиса [<http://ruscorpora.ru/poetic>] (<http://ruscorpora.ru/poetic>)

4. Poetry Foundation API и Corpus

Представляет доступ к поэтическим текстам современных и классических авторов; удобен для анализа тем, мотивов, жанров.

5. Oxford Text Archive

Это коллекция размеченных текстов, включая поэзию на английском языке различных эпох, с богатой аннотацией.

В настоящее время в рамках проекта «Развитие казахского поэтического корпуса: морфологическое и поэтическое обозначение стихотворений Абая» (грант № BR21882334)

разработан поэтический корпус (<http://corpora.kz/>). Объектом исследования является казахский малый поэтический корпус, сфокусированный на стихах Абая Кунанбаева, выдающегося казахского поэта и мыслителя. Исследование включает морфологическую и поэтическую разметку его произведений, а также разработку цифровой базы данных для хранения и обработки этой информации. Целью работы является создание структурированного и всесторонне аннотированного малого поэтического корпуса стихотворений Абая, который будет способствовать лучшему пониманию и анализу его творчества с точки зрения морфологии и поэтической структуры.

Разметка – один из важнейших аспектов любого корпуса, в том числе и поэтического корпуса. Например, национальные корпуса имеют:

- морфологическую и синтаксическую разметку;
- лексико-семантическую аннотацию;
- авторство и дату написания.

Однако поэтический корпус различается тем, что помимо приведенных разметок он включает:

- метрическую разметку (ямб, хорей, дактиль и др.);
- сегментацию по строфам и строкам;
- рифменную схему (АВAB, АABВ и т.д.);
- белый стих и т.п.

Трудности и перспективы.

Создание поэтических корпусов связано с рядом трудностей:

- автоматическое распознавание метра и рифмы требует сложных алгоритмов;
- разметка часто проводится вручную, что увеличивает стоимость и длительность проектов;
- поэтический язык нестабилен, поскольку существуют различные лексические и грамматические особенности: наличие архаизмов, неологизмов, окказионализмов, нарушения грамматики.

Тем не менее, развитие нейросетей и методов NLP (Natural Language Processing) открывает новые возможности для создания как поэтических корпусов, так и Национальных корпусов в целом: автоматизация разметки, генерация текстов, в том числе поэзии, стилистическая атрибуция и прочее.

Благодарности.

Данное исследование финансируется Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант № BR21882334 «Развитие казахского поэтического корпуса: морфологическое и поэтическое обозначение стихотворений Абая»).

Использованная литература:

1. <https://kartaslov.ru/>
2. <https://www.nur.kz/leisure/interesting-facts/1553089-skolko-yazykov-sushchestvuet-v-mire/>
3. Ethnologue: Languages of the World. <https://www.ethnologue.com/>
4. Жубанов А.К., Жанабекова А.А. Корпусная лингвистика. – Алматы: Қазақ тілі, 2016. – 336 с.
5. Национальный корпус казахского языка [Электрон. ресурс]. <https://qazcorpus.kz/about/1/>
6. Малые корпуса национального корпуса казахского языка [Электрон. ресурс]. <https://qazcorpora.kz/search>
7. Момынова Б., Имангазина М., Анесова У. Разработка лексико-семантической разметки глаголов в национальном корпусе казахского языка: мировой опыт, классификация, разметка в корпусе // КУМОиМЯ им. Абылайхана, фил.сер. – 2022. – Т. 66. №3. – С. 128-146.

BADIIY ASARLAR PARALLEL KORPUSI UCHUN SIFAT SO‘Z TURKUMIGA OID FRAZEOLOGIK BIRLIKLAR BAZASINI YARATISH HAQIDA

Sirojiddinova Muxarram Rahimjon qizi

Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti
Kompyuter lingvistikasi va amaliy tilshunoslik
kafedrası 2-bosqich magistranti

E-mail: muhammadxonsirojiddinova@gmail.com

Annotatsiya: Maqolada ingliz va o‘zbek tillaridagi frazemlar kontekstologik metod asosida tahlil qilindi, ularning umumiy va farqli jihatlari nazariy jihatdan asoslandi. Asarlardagi muqobil frazeologik birliklar Excel dasturida segmentlandi. Frazemalar [adjective +noun], [adjective +and +adjectiva], [preposition + adjective], [adjectiva + preposition + noun], [as + adjectiva + as + noun or adjectiva + as + noun] tarzida morfologik teglandi.

Kalit so‘zlar: parallel korpus, segmentlash, kontekstologik metod, sifat so‘z turkumiga oid frazeologik birliklar, morfologik teglash.

Abstract: The article analyzes phraseological units in English and Uzbek using the contextual method, and their common and different aspects are theoretically substantiated. Alternative phraseological units in the works were segmented in Excel. Phraseological units were morphologically tagged in the following ways: [adjective +noun], [adjective +and +adjectiva], [preposition + adjective], [adjectiva + preposition + noun], [as + adjectiva + as + noun or adjectiva + as + noun].

Keywords: parallel corpus, segmentation, contextual method, phraseological units of the adjective word class, morphological tagging.

Аннотация: В статье анализируются идиомы английского и узбекского языков с использованием контекстуального метода, теоретически обосновываются их общие и различные стороны. Альтернативные фразеологические единицы в работах сегментировались в Excel. Фразы были морфологически помечены в форме [прилагательное + существительное], [прилагательное + и + прилагательное], [предлог + прилагательное], [прилагательное + предлог + существительное], [как + прилагательное + как + существительное или прилагательное + как + существительное].

Ключевые слова: параллельный корпус, сегментация, контекстуальный метод, фразеологизмы класса прилагательных, морфологическая разметка.

Frazeologik birliklarni morfologik tahlil asosida o‘rgangan A.V. Kunin sifat komponentli frazeologik birliklarga ham alohida e‘tibor qaratgan. Shuningdek, R.Moon, J.C.Cowie, J.R.Firth, P.H.Peters, G.Makkai, F.Smith kabi olimlar ingliz tilidagi idioma, kollokatsiya va boshqa frazeologik birliklarni o‘rganish asosida sifat komponentli frazeologik birliklarning semantik, funksional xususiyatlarini tahlil qilganlar.

O‘zbek tilshunosligida ingliz-o‘zbek tillaridagi sifat komponentli frazeologik birliklarni tadqiq etish bo‘yicha tizimli ilmiy ish hali yaratilmagan bo‘lsa-da, M.I.Xolmatova, Z.Narmuratov kabi tadqiqotchilar bu masalasiga to‘xtalib o‘tishgan. Xususan, M.I.Xolmatova ingliz-o‘zbek tillaridagi sifat komponentli iboralarni qiyosiy-tipologik, lingvokulturologik, morfologik jihatdan tahlil etgan bir nechta maqolalari chop etgan. Ular “Ingliz-o‘zbek tillaridagi sifat komponentli iboralarning lingvokulturologik qiyosiy tahlili” nomli maqolasida sifat komponentli iboralarni grammatik tuzilishiga ko‘ra 5 guruhga ajratgan. [Xolmatova, 2021;723].

Ingliz yozuvchilar o‘z asarlarida ishlatgan frazeologik birliklarni kontekstologik metod asosida tahlil qildik. Natijalar shuni ko‘rsatadiki, Ernest Xemenguey, Mark Tven va boshqa

yozuvchilar badiiy asarlarda frazeologik birliklarni o'z o'rnida qo'llay bilgan. Asarlarda fe'l va ot komponentli birliklar ko'p qismni tashkil qilgan, sifat komponentli frazeologik birliklar esa buning aksi. Shunday bo'lishiga qaramasdan to'pangan bazadagisifat komponentli frazemalarni morfologik tamoyilga ko'ra quyidagi guruhlariga ajratib tasnifladik:

- [Adjective +noun]
- [Adjective +and +adjectiva]
- [Preposition + adjective]
- [Adjectiva + preposition + noun]
- [As + adjectiva + as + noun or adjectiva + as + noun]

1. [Adjective-noun] – qolipidagi frazeologik birliklar abstrakt tushunchalarni kuchaytirish uchun ishlatilib, metaforik yoki timsolli ma'no kasb etishi mumkin. Bunda sifat darajaviy belgisiga ko'ra oddiy, qiyosiy va orttirma darajaga egaligi bilan ahamiyatlidir.

Big mouth. O'zbek tilida: *og'zini qopqog'i yo'q, gullab qo'yadigan*

Big mouth – ushbu frazema so'zma-so'z "**katta og'iz**" deb tarjima qilinadi. O'zbek tilida "**katta og'iz**" iborasi doim *katta gapiradigan, dimog'dor, maqtanchoq, kibrli insonlarga* nisbatan ishlatiladi. Inglizlar **big mouth** iborasini *og'zida gap turmaydigan, gullab qo'yadigan, maqtanchoq insonlarga* nisbatan qo'llaydi. Ko'rib turganingizdek, inglizlar ishlatadigan "**big mouth**" iborasi o'zbeklar ishlatadigan "**katta og'iz**" iborasiga ma'no jihatdan deyarli o'xshash.

Thick skin. O'zbek tilida: *Terisi qalin.*

Thick skin – o'zbek tilida "**terisi qalin**" iborasi "*gap kor qilmaydigan, beti qattiq*" degan ma'noni anglatib, birovning gapini pisand qilmaydigan insonlarga nisbatan ishlatiladi. Xayratlanarlisi, inglizlar ham mana shundan beti qattiq insonlarga nisbatan *thick shin* idiomasi qo'llaydilar.

Thin skin. O'zbek tilida: *Terisi yupqa.*

Thin skin – ushbu idioma **thick skin** idiomasi nisbatan antonimlik xosil qilib, "**terisi yuqa**" deb tarjima qilinib, *gap ko'tara olmaydigan, ko'ngli nozik insonlarga* nisbatan qo'llaniladi.

For example: *If you have thin skin, you will never survive in real life.*

Tarjima: *Agar sen shunaqa tering yupqa bo'lsa, hayotda tirik qola olmayсан.*

T/F	Frazema	Definition	Tarjimasi	Example
1	Fast talker	A person who speaks quickly, often to persuade or deceive others	sergap, ko'pgap	He was a fast talker, always convincing others to take risks they shouldn't.
2	Fast worker	A person who works quickly and efficiently	harakati tez	She was fast worker finishing all her tasks before the deadline.
3	(Have) a loose tongue	To talk too much or reveal secrets carelessly	tilida gap turmaydigan	He had a loose tongue and soon the entire town knew about the plans.
4	(Have) a quick temper	Too become angry easily	jahli tez	She had quick temper and often lost control over small problems.
5	White Lie	A small harmless lie.	beozor yolg'on	"Scout told a white lie to keep her friend out of trouble. (From To Kill a Mockingbird by Harper Lee)

2. [Adjective +and +adjectiva] qolipidagi idiomalar ingliz tilida sifat komponentli idiomalarning katta qismini tashkil qiladi. **And** bog'lovchisi yordamida bog'langan sifatlar, asosan, juft sifat bo'lib, *ma'nosi bir-biriga yaqin yoki qarama-qarshi ma'noga* ega hisoblanadi. Ikki juft sifat o'zaro bog'lanib ma'noni yanada kuchaytirgan holda yana yangi ma'no hosil qiladi.

High and mighty. O'zbek tilida: *Dimog'dor, kerilgan.*

High and mighty – acting superior, arrogant, or self-important. Ushbu idioma so‘zma-so‘z baland va kuchli deb tarjima qilinadi. Bu ikki juft sifat ma‘no jihatdan o‘zaro bir-biriga yaqin bo‘lib, birgalikda “*dimog‘dor, kerik*” degan ma‘noda qo‘llaniladi.

For example: “Scarlett found herself resenting Ashley’s high and mighty attitude”. (*Gone with the wind* by Margaret Mitchell)

Tarjima: Skarlet Eshlining kerik munosabatidan ranjidi.

Black and White. O‘zbek tilida : Aniq va ravshan, yaxshi va yomon tomoni.

Black and White – A clear distinction between right and wrong, without complexity. Ushbu idioma bir-biriga zid bo‘lgan ranglarni anglatuvchi sifatlardan tuzilgan bo‘lib, so‘zma-so‘z “*qora va oq*” deb tarjima qilinadi. Idioma ikki zid ma‘noli sifatlarning birikishidan tashlik topgan. Inglizlar hammasi aniq va tushunarli, qo‘shimcha talqinlarga o‘rin yo‘q bo‘lgan vaziyatlarga nisbatan qo‘llaydi. Shuningdek, ingliz va o‘zbek tillarida oq rang yaxshilikni, qora rang esa yomonlikning ramzi sifatida qaraladi. Shu sababli, idioma biror narsaning yaxshi va yomoni tomonini ifodalash uchun ham ishlatiladi.

For example: “Atticus knew that justice was not always black and white”¹²⁵.

Tarjima: “Attikus adolat har doim ham aniq va ravshan emasligini bilardi”.

T/R	Frazema	Definition	Tarjimasi	Example
1	Pure and Simple	Absolutely clear, without any hidden meaning or complexity.	shubhasiz, -dan boshqa narsa emas	"Elizabeth rejected Mr. Collins' proposal, pure and simple". (From <i>Pride and Prejudice</i> by Jane Austen)
2	Bright and Early	Very early in the morning.	ertalab barvaqt	"Gatsby arrived bright and early, hoping to see Daisy". (From <i>The Great Gatsby</i> by F. Scott Fitzgerald)
3	Free and Clear	Without any obstacles or obligations.	yengil, yoqimli	"Mitch wanted to leave the company free and clear, without any debts". (From <i>The Firm</i> by John Grisham)
4	Rough and Ready	Basic but effective.	taxminiy qilingan, qo‘pol	"The pirate's plan was rough and ready, but it worked". (From <i>Treasure Island</i> by Robert Louis Stevenson)
5	Fine and Dandy	Everything is okay, often used sarcastically.	ajoyib va zo‘r	"Everything was just fine and dandy, except it wasn't". (From <i>The Catcher in the Rye</i> by J.D. Salinger)

3. [Preposition + adjective] ushbu qolibdagi frazemalar ingliz tilida unchalik ko‘p emas va ularning aksariyati kirish so‘z sifatida qo‘llaniladi.

At worst. O‘zbek tilida : nari borsa, eng yomoni.

At worst – in the most negative or unfavorable scenario. Ushbu birikma kirish so‘z sifatida qo‘llanilib, gapning negative ma‘noga ega ekanligini anglatib turadi.

For example: “At worst, they would be caught, but they were willing to take the risk.”

Tarjima: Eng yomoni, ular ushlanib qolishi mumkin edi, lekin ular tavakkal qilishni xohladi.

In short. O‘zbek tilida : qisqasi.

In short – in summary, briefly. Ushbu birikma kirish so‘z sifatida qo‘llanilib, gapning yakuniy qismi ekanligini anglatib turadi.

For example: In short, the expedition was a failure, but they learned valuable lesson.

Tarjima: Qisqasi, ekspeditsiya muvaffaqiyatsiz bo‘ldi, lekin ular qimmatli saboq oldi.

4. [Adjectiva + preposition + noun].

Full of beans. O‘zbek tilida: serg‘ayrat, shijoatga to‘la.

Full of beans – very energetic and lively. Ushbu ibora soʻzma-soʻz *“loviyaga toʻla”* deb tarjima qilinadi. Loviya yetishtirish koʻp mehnat talab qilganligi bois inglizlar koʻp mehnat qiladigan, serharakat insonlarga nisbatan ushbu birikani qoʻllaydi. *“Ful of beans”* iborasi metanimiya asosida sergʻayrat insonlarga oʻxshatilgan.

For example: *"Anne was full of beans, always ready for a new adventure".*

Tarjima: *"Anne doimo yangi sarguzashtlarga tayyor boʻlib, juda sergʻayrat edi."*

Sick at heart. Oʻzbek tilida: *ich-ichidan hafa.*

Sick at heart – Feeling deeply sad or distressed. Ushbu idiomasi soʻzma-soʻz *“yuragi kasal”* deb tarjima qilinadi. Oʻzbek tilida *“yuragi kasal”* frazemasi *“kimdandir qattiq ranjigan”* yoki *“muammosi bor”* degan maʼnoni anglatadi. Yuragingizga yaqin inson sizni qattiq ranjitsa yoki biror muammoga duch kelsa, yurakda ogʻriq paydo boʻladi. Shu sababli inglizlar *“sick at heart”* idiomasi qattiq depressiyaga tushgan, juda xafa boʻlgan insonlarga nisbatan qoʻllaydi.

For example: *David was sick at heart about having to leave his family behind.*

Tarjima: *David oilasini tark etishidan juda qattiq xafa edi.*

5. [As + adjectiva + as + noun or adjectiva + as + noun] qolipidagi idiomalar ingliz tilida oʻzbek tiliga nisbatan ham koʻproq ekanligi tadqiqot natijasida maʼlum boʻldi. *As... as (day, dek, kabi) bogʻlovchisi* yordamida bogʻlangan soʻzlar koʻchma maʼno hisobiga oʻz maʼnolarini qisman yoʻqotib, turgʻun birikmaga aylangan. Teng bogʻlangan qiyosiy iboralarda adjectivedan oldin kelga *as* bogʻlovchisi qoʻllanilishi yoki tushib qolishi mumkin, lekin ot (noun) oldidagi *as* bogʻlovchisi doim boʻlishi shart.

T/F	Frazema	Definition	Tarjimasi	Example
1	As brave as a lion	Extremely courageous and fearless.	sherdek jasur	"Lucy felt as brave as a lion when she stood up to the White Witch ". (From The Chronicles of Narnia by C.S. Lewis)
2	As cold as ice	Emotionally distant or unfeeling.	muzdek sovuq, hissiz	"Heathcliff's stare was as cold as ice after Catherine betrayed him". (From Wuthering Heights by Emily Brontë)
3	As light as a feather	Very light in weight.	patdek yengil	"Wendy felt as light as a feather when she flew with Peter Pan". (From Peter Pan by J.M. Barrie)
4	As wise as an owl	Very intelligent or knowledgeable.	boyqushdek dono	"Gandalf was as wise as an owl, always giving the best advice". (From The Hobbit by J.R.R. Tolkien)
5	As blind as a bat	Unable to see well or completely unaware.	koʻrshapal akdek koʻr	"The townspeople were as blind as a bat to the injustice around them". (From To Kill a Mockingbird by Harper Lee).

Badiiy asarlarda qoʻllanilgan frazeologik birliklar kontekstologik metod asosida izohlanildi va morfologik tamoyilga koʻra ot, sifat, feʼl komponentli birikmlarga ajratib tasniflash maʼqul koʻrildi. Maqolada ulardan sifat komponentli frazeologik birliklar ajratilib, qiyosiy-tipologik, semantik, morfologik jihatdan tahlil qilindi. Tahlil shuni koʻrsatadiki, turli til oilalariga mansub boʻlgan ingliz va oʻzbek tillari til qurilishi, leksik birliklar va ularning grammatik xususiyatlari bir-biridan keskin farqlanishiga qaramay, sifat komponentli frazemalarning maʼnolari bir-biriga qisman yaqin. Ular ikkala tilda ham insonning fazilat yoki qusurlarini ifodalashga xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Abdurahmonova N. English-uzbek dictionary of phrasal verbs. – Toshkent.: Turon zamin ziyo, 2017.
2. Abdurakhmonova, N. (2016). The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta, 2(38), 12-7.
3. Suleymanov, D., Nevzorova, O., Gatiatullin, A., Gilmullin, R., & Khakimov, B. (2013). National corpus of the Tatar language “Tugan Tel”: grammatical annotation and implementation. Procedia-Social and Behavioral Sciences, 95, 68-74.
4. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: [uzbekcorpus. uz](http://uzbekcorpus.uz). In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.
5. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021). Linguistic functionality of Uzbek Electron Corpus: [Uzbekcorpus. uz](http://Uzbekcorpus.uz). In International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities, ICISCT 2021.
6. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzworonet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
7. Xolibekova O. K. Ingliz va o‘zbek tillarida zoonim va fitonim komponentli frazeologik birliklarning semantikstruktur, lingvomadaniy tadqiqi. Filol. fan. b. fals. dok. (Phd) diss. avtoref. –Samarqand, – 2023. –131 b.
8. Xolmatova M.I.. Ingliz-o‘zbek tillaridagi sifat komponentli iboralarning lingvokulturologik qiyosiy tahlili. Scientific Journal Impact Factor (SJIF) 2021:5. –723bet.
9. Z.Narmuratov “Ta’lim” va “Ilm” mazmunli frazemalarning semantikstruktur, lingvomadaniy xususiyatlari (Ingliz va o‘zbek tillari misolida). Filol. fan. dok. (Dsc) diss. avtoref. –Buxoro, 2024. – 246 b.
10. Mark Tver. The Adventure of Tom Sawyer.
11. Mark Tven. Tom Soyerning boshidan kechirganlari. (Ilyos Muslim tarjimasi). – Toshkent.: Yangi asr avlodi, 2014.

KORPUS UCHUN DIALEKTAL LINGVISTIK BAZA YARATISH: MUAMMO VA YECHIMLAR

Abduraxmonova N.Z.

Filologiya fanlari doktori, professor

Abdullayeva N. M.

O'zMU Kompyuter lingvistikasi
mutaxassisligi 1- kurs magistranti

E-mail: nasimaabdulla35@gmail.com

ORCID-0009-0000-8285-9369

Annotatsiya: Maqolada korpus uchun dialektal lingvistik baza yig'ish, uning dolzarbligi va bugungi kundagi ahamiyati, dialektal baza yaratishdagi muammolar va ularning yechimlari atroflicha tahlil qilinadi, Qarshi shahri lingvistik bazasini yaratish bo'yicha ba'zi namunalar yoritiladi.

Kalit so'zlar: dialektal lingvistik baza, korpus, transkripsiya, lingvistik teglash, metama'lumot.

Annotation: The article provides a detailed analysis of the collection of a dialectal linguistic database for a corpus, its relevance and importance today, as well as the problems encountered in the creation of such a database and their possible solutions. It also presents some examples related to the development of a linguistic database for the city of Qarshi.

Keywords: dialectal linguistic database, corpus, transcription, linguistic tagging, metadata

Аннотация: В статье подробно анализируются сбор диалектологической лингвистической базы для корпуса, её актуальность и значение на сегодняшний день, проблемы при создании диалектной базы и пути их решения. Также освещаются некоторые примеры по созданию лингвистической базы города Карши.

Ключевые слова: диалектологическая лингвистическая база, корпус, транскрипция, лингвистическая разметка, метаданные

Bugungi globallashayotgan jamiyatda kompyuter lingvistikasi va korpus tilshunosligi tez sur'atlar bilan rivojlanib, tabiiy tilni qayta ishlash, avtomatik tarjima, sun'iy intellektning rivojlantiruvchi ma'lumotlar bazasi shakllantirishda muhim ahamiyat kasb etib kelmoqda. Bunday baza shakllantirish ko'p bosqichli murakkab jarayon bo'lib, lingvistik ma'lumotlarni tizimli yig'ish, saqlash va tahlil qilishga xizmat qiluvchi elektron ma'lumotlar ombori hosil qilishda tilshunoslikning har bir bo'g'ini chuqur tahlilga tortilgan holatda korpusdan joy olmoq'i lozim. Xususan dialektal ma'lumotlar ham. Dialektal lingvistik baza lahja va shevalarga oid fonetik, morfologik, leksik, semantik ma'lumotlarni o'z ichiga olishi zarur va ushbu bazalar til o'zgarishlarini kuzatish, tarixiy tilshunoslikni rivojlantirish va tilni raqamlashtirish uchun xizmat qilmog'i darkor. Bunday bazalardan shakllangan dialektal korpuslar so'nggi davrlarda tadqiqi sustlashgan dialektal soha rivojini ham yuqori pog'onaga olib chiqa oladi. Yurtimizning har bir hududi yuzasidan amalga oshirilgan dialektal tadqiqotlarning korpusda aks etishi nafaqat til taraqqiyotiga, balki zamonaviy raqamli texnologiyalar yaratishda muhim lingvistik asos bo'lib xizmat qiladi.

Dunyodagi ko'plab ilm-fan rivojlangan mamlakatlarning korpusshunoslik tadqiqotlariga nazar tashlasak, dialektal bazaga ham keng urg'u berilganini ko'ramiz. Xususan, Britaniyaning Xelsinki korpusi, Rus tili korpusidan joy olgan dialektal subkorpus, Hindiston dialektlari korpusi, Xitoy Mandarin dialektlari korpusi, arab dialektlari sotsial Mediya korpusi kabi. Ularda sheva vakillarining audio va turli yozma manbalardan to'plangan va mukammal lingvistik hamda

ekstralingvistik tahlilga tortilgan turli matnlardan tashkil topganini ko‘rishimiz mumkin. So‘nggi yillarda yurtimizda ham dunyo korpusshunoslik metodologiyasiga asoslangan holda bir qator dialektal tadqiqotlarimiz maydoni kengaymoqda. Xususan, tadqiqotchi M. Xolova Boysun tuman shevasi asosida ilk bora “o‘zbek milliy shevalar korpusi” (<https://dialect.uz/>)ni tuzdi. Qidiruv amalga oshirish uchun belgilangan maydonlarda adabiy til, transkripsiya, transliteratsiya, turkum, dialekt etimologiyasi, dialektni ichiga olgan gap tarkibi, sinonim, omonim va antonimligi hamda dialekt so‘zning rus va ingliz tilidagi tarjimalari aks etadi. Tadqiqotchi korpus tuzishni jahondagi mashhur dialektal korpuslar metodologiyasiga asoslangan holatda ikki bosqichda amalga oshirish kerakligini ta’kidlaydi va bu bosqichlarning biri sifatida korpusning texnik va ma’lumotlar bazasini shakllantirish, ya’ni dialektal korpusning pasporti, mavzuga oid guruhlari, matn va janrlar bo‘yicha metarazmetkasini tuzishni hamda ikkinchi bosqich sifatida to‘plangan ma’lumotlar asosida fonetik va morfologik metarazmetkasini ishlab chiqish lozimligini ko‘rsatadi [Холова М. 2022]. Denov tumani shevalari yuzasidan tadqiqotchi A. Turdialiyevning hudud shevalarining asosiy xususiyatlari, ularning bir-biriga ta’siri tadqiq qilingan va ma’lumotlar bazasi shakllantirilib “o‘zbek tili korpusi”ning dialektologik subkorpusiga joylashtirilgan tadqiqot ishi ham o‘zbek dialektal korpusi boyitilishida muhim qadam bo‘ldi. [Turdialiyev A. 2023]. Bularidan tashqari o‘zbek tili ta’limiy korpusi, o‘zbek tili mualliflik korpusi kabi bir qator manbalarimizda ham o‘zbek tilining asosiy 3 lahjasi bo‘lmish, qarluq, qipchoq, o‘g‘uz lahjalariga oid adabiy janrlar, ijtimoiy-siyosiy materiallardan namunalar beriladi va dialektal subkorpusni muntazam boyitib turish maqsad qilinadi. O‘zbek tili elektron korpusiga og‘zaki nutqlarni joylashtirish masalalarida ham dialektal audionutqlarni metama’lumotlarini alohida qayd etgan holatda korpusga joylash masalasiga e’tibor qaratiladi. Og‘zaki nutqning etnografik ma’lumotlarni metama’lumot sifatida berish ham muayyan darajada asosiy talablardan biri. So‘zlovchining qaysi hududga tegishli ekanligi, jinsi, yoshi, dialekti, kasbi bularning hammasi audiomatnni korpusga kiritishda bosh mezon sanaladi [file:///C:/Users/user/Downloads/o-zbek-tili-elektron-korpusida-http-uzbekcorpus-uz-og-zaki-matnlar-korpusini-yaratishning-nazariy-va-amaliy-masalalari%20(1).pdf] .

Korpus uchun dunyo dialektal korpusshunoslik an’analariga muvofiq lingvistik baza yaratish yurtimizning deyarli har bir hududi areal tadqiq uchun muhtoj bo‘lgan bugungi kun uchun ancha murakkablik tug‘diradi. Dialektal baza shakllantirishdagi asosiy muammolardan biri materiallar yetishmasligi. Ko‘pgina shevalar faqat og‘zaki nutqda mavjud bo‘lib, yozma shakli mavjud emas. Bu esa ularni hujjatlashtirish va raqamli formatga o‘tkazishda qiyinchilik tug‘diradi [J. Milroy, 1992]. Dialektal farqlarning notekisligi ham jarayonni murakkablashtiradi. Biror hudud ichida ham bir necha kichik dialektlar mavjud bo‘lishi mumkin. Bu esa umumiy va universal bazani yaratishni qiyinlashtiradi [Trudgill, 2004]. Dialektal matnlarni lingvistik va ekstralingvistik razmetkalashning qo‘lda bajarilishi esa dialektal korpuslarni tuzishda uzoq vaqtning talab qiladi. Sheva so‘zlari ko‘pincha rasmiy tildan farq qilgani sababli avtomatik morfologik analizatorlar ularni noto‘g‘ri qayta ishlash mumkin [Jurafski & Martin, 2023].

O‘zbek korpusida dialektal bazani yaratish sanab o‘tilgan bir qator qiyinchiliklarga qaramay kechiktirilishi mumkin bo‘lmagan muhim tadqiqot yo‘nalishi sifatida O‘zbekiston milliy universiteti Kompyuter lingvistikasi kafedrasida mutaxassislar hamda talabalari tomonidan yurtimizning har bir hududi bo‘yicha korpus uchun ma’lumotlar yig‘ilib lingvistik tahlillar olib borish uchun izchil tadqiqotlar amalga oshirilmoqda. Xususan, Qashqadaryo viloyatining Qarshi shahri shevasi ham ilk bor korpus uchun lingvistik baza sifatida tadqiqi olib borilmoqda. Hudud shevalari bir qator dialektolog olimlarimiz tomonidan (Y.D. Polivanov, V.V. Reshetov, A.Shermatov va boshq.) o‘rganib kelingan bo‘lsada, korpus uchun lingvistik annotatsiyalangan ma’lumotlar bazasi shakllantirish alohida tadqiqotni talab qiladi. Qarshi shahri dialekti Qarluq-chigil-uyg‘ur lahjasi tarkibiga kiritilib Qarshi gruppasiga Qarshi, Shahrisabz, Kitob, Yakkabog‘ va (qipchoq shevalaridan tashqari) Qashqadaryodagi shu tipdagi shevalar kiradi (V.V.Reshetov, 1966). Qarshi shahri o‘zbek tilining y-lovchi shevalari gruppasiga kiradi. Bu shevada unlilar garmoniyasi yo‘q. Mazkur sheva o‘zining grammatik va leksik xususiyatlari bilan o‘zbek tilining boshqa y-lovchi shevalariga o‘xshaydi, shuningdek, ulardan farq qiladigan tomonlarga ham

egadir(V.V. Reshetov II kitob, 178-b). Hudud dialekti yuzasidan lingvistik baza shakllantirishda dastavval shaharning tub aholisi joy olgan eski mahallalarida istiqomat qiluvchi vakillaridan audionutqlar yozib olinmoqda hamda metama'lumoti shakllantirilmoqda. Shahar shevasining atrofda tumanlarga ta'sir doirasini o'rganish maqsadida hamda atrof tumanlardagi lahjalarning shahar aholisi nutqiga bevosita aloqasini qiyoslash maqsadida shahar atrofida tumanlarda yashovchi aholidan ham audionutqlar yozib olinishi rejalashtirilganki, bu orqali hudud lahjasi va chegara hududlarga tarqalgan oraliq shevalar aniqlanadi. Yozib olingan audionutq dastavval lotin tiliga asoslangan yangi o'zbek transkripsiyasiga asoslanib matnga o'giriladi hamda shevaga xos bo'lgan so'zlar aniqlanib, adabiy til bilan muqobili keltiriladi, izohlari beriladi hamda grammatik teglanadi. So'ngra matn metama'lumoti shakllantiriladi. Quyidagi jadvalda shevaga oid to'plangan ma'lumotlar adabiy til, transkripsiya, lahja va metama'lumotlari bilan keltirilgan.

t/r	Adabiy tilda	Shevada (talaffuz i)	Transkripsiya	Lahjasi	Metama'lumoti			
					Yashash manzili	Yoshi va jinsi	Ma'lumot i, kasbi	Uru g'i
1	Singlimni	Singlimni	Singlimni	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
2	nevara	nevara	nevärä	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
3	ko'rgandi	ko'ruvdi	koruvdi	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
4	shuni	shuni	šuni	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
5	beshikka	beshikka	bešikkä	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
6	belashdi	belashdi	belašdi	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
7	O'sha yoqqa	o'sheqqa	Ošeqqä	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
8	bordik-da	bordiyda	bärdijda	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
9	qudalari	qudalari	qudäläri	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
10	keldi	keldi	keldi	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
11	ularni	ulani	uläni	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
12	jo'natdik	jo'natdik	žonätdik	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
13	mana	mana	mänä	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
14	shu	shuni	šu	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
15	bizlarda ham	bizadaya m	bizädäjä m	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
16	mana	mana	mänä	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
17	fotiha	potiya	pätijä	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	
18	to'yi	to'yi	toji	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	

19	bo'layapti	bo'lopti	bolāpti	Qarluq	Qarshi shahri,	70 yosh, ayol	o'rta, hamshira	

Dialekt matnlarni lingvistik teglash ham qo'lda amalga oshiriladigan jarayondir. O'zbek tilida taklif qilinayotgan avtomatik teglash tadqiqotlari ham bugungi kunda mavjud. Professor N. Abduraxmonova tomonidan uzbekcorpus.uz dasturida o'zbek tilida ilk bor FST (finite state transducer), ya'ni, chekli avtomat transyuteri vositasida korpusni morfologik teglashning avtomatik usuli ishlab chiqilgan va amaliy ifodasi sifatida <http://uzbekcorpus.uz> saytining qidiruv tizimida lemma, token hamda *n-gram* modeli asosida konkordanslarning qidiruv menejeri yaratilgan. Boshqa bir qator tilshunos olimlarimiz tarafidan ham avtomatik tegsetlar bo'yicha taklif qilingan tadqiqotlar mavjud (Elov B., Hamroyeva SH., Abdullayeva O., Uzoqova M. 2022. "o'zbek tilida PoS tegging masalasi: muammo va takliflar". *o'zbekiston: til va madaniyat. Amaliy filologiya*. 2 (5): 51-68). Ammo shevaga xos so'z standart qolipga mos kelmagani uchun qo'lda teglab chiqiladi. Masalan, jadvalda keltirilgan ba'zi shevaga oid so'zlar quyidagicha grammatik teglanadi:

1. *Bordiyda*: V+Past+Pl[1P]+Postp
 Bu yerda Bor V(fe'l)+di Past(o'tgan zamon)+y(-k 1-shaxs ko'plik qo'shimchasining sheva shakli) Pl[1P](1-shaxs ko'plik)+da Postp(yuklama).
2. *Bizadayam*: Pron+Pl+Loc+Foc
 Ushbu so'z bizlarda ham so'ziga to'g'ri kelgani uchun Biz Pron(olmosh)+lar Pl(ko'plik qo'shimchasi)+da Loc(o'rin-payt kelishigi)+ham Focus(ya'ni ta'kidlovchi birlik).
3. *Potiya*: N+Sg+Nom
 Fotiha so'zi N(ot)+Sg(birlikda)+Nom(nominativ, asos)
4. *Bo'lopti*: V+Pres+Prog
 Bo'lyapti so'zi Bo'l V(bo'lmoq fe'li)+opti(-yapti, hozirgi zamon davom fe'li)Pres+Prog(hozirgi zamon davomiylik).
 Shu tartibda kelgusidagi ishlarimizda hudud shevasining yanada o'ziga xos jihatga ega so'zlari aks etgan audionutqlar to'planib transkripsiyalangan va teglangan holatda metama'lumotlari ila korpus uchun baza yaratish uchun tadqiq qilinishi ko'zda tutilgan. Yurtimizning har bir hududi dialektlari mana shu tamoyillar asosida tadqiq etilib dialektal baza yaratish esa istiqbolini kutib turibdi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N.Z. Korpus lingvistikasi. Darslik. GlobeEdit, 2023
2. Abdurakhmonova, N. (2016). The bases of automatic morphological analysis for machine translation. *Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta*, 2(38), 12-7.
3. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus.uz. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.
4. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021). Linguistic functionality of Uzbek Electron Corpus: [Uzbekcorpus.uz](http://uzbekcorpus.uz). In International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities, ICISCT 2021.
5. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
6. Turdaliyev A.X. Denov shevalarini areal o'rganish va til korpusiga joylashtirish. *Filologiya*

7. bo'yicha falsafa doktori (PhD) ilmiy darajasini olish uchun yozilgan dissertatsiya. T.: 2023, 129-b
8. Холова М. Ўзбек миллий шевалари корпуси тадқиқи (Бойсун тумани “ж”ловчи шевалари мисолида). Монография, – Термиз: ТерДУ НММ, 2022.
9. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
10. Elov B., Hamroyeva SH., Abdullayeva O., Uzoqova M. 2022. “o‘zbek tilida PoS tagging masalasi: muammo va takliflar”. o‘zbekiston: til va madaniyat. Amaliy filologiya. 2 (5): 51-68
11. Jurafsky, D., & Martin, J. H. Speech and Language Processing. Pearson, 2023
12. Milroy, J. Linguistic variation and change: On the Historical Sociolinguistics of English. Blackwell, 1992
13. Trudgill, P. New-Dialect Formation: The inevitability of Colonial Englishes. Edinburgh University Press, 2004
14. file:///C:/Users/user/Downloads/o-zbek-tili-elektron-korpusida-http-uzbekcorpus-uz-og-zaki-matnlar-korpusini-yaratishning-nazariy-va-amaliy masalalari%20(1).pdf
15. <http://uzbekcorpus.uz>
16. <https://uzschoolcorpara.uz/>

FONETIK VA LEKSIK JIHATIDAN BUXORO SHEVASINING TIPIK XUSUSIYATLARI

Soliyeva Farida Bobir qizi

O'zbekiston Milliy universiteti Kompyuter lingvistikasi
va Amaliy tilshunoslik ixtisosligi 1-kurs magistranti

E-mail: faridasoliyeva111@gmail.com

Annotatsiya: Keltirilayotgan ma'lumotlarda Buxoro viloyatining Vobkent tumanida istiqomat qiluvchi aholining kundalik turmush tarzida qo'llaydigan ayrim dialektal so'zlarning tasnifi yoritilgan. Shevaga oid so'zlarning etimologiyasi, fonetikasi hamda leksikasi o'rganilib, adabiy tilda va shevada qiyosiy tahlili ko'rib chiqilgan.

Tayanch so'zlar: sheva, adabiy til, fonetika, leksika, ma'noviy guruh, etimologiya, qiyosiy tahlil.

Annotation: The provided information discusses the classification of certain dialectal words used in the daily life of the population living in the Vobkent district of Bukhara region. The etymology, phonetics, and lexicon of the dialect are studied, and a comparative analysis of these words in both the literary language and the dialect is presented.

Keywords: Phoneme, literary language, phonetics, lexicon, semantic group, etymology, comparative analysis.

Аннотация: В предоставленной информации освещена классификация некоторых диалектных слов, используемых в повседневной жизни жителей Вобкентского района Бухарской области. Изучена этимология, фонетика и лексика слов, относящихся к говору, а также проведён сравнительный анализ между литературным языком и диалектом.

Ключевые слова: Говор, литературный язык, фонетика, лексика, семантическая группа, этимология, сравнительный анализ.

Kirish

Shevashunosligimizning tarixi XI asrning mashhur tilshunosi Mahmud Koshg'ariyga borib taqaladi. Uning "Devonu lug'atit turk" asari turkiy xalqlar shevalarining so'zligini arab tilida izohlab, o'zida jamlagan mukammal va bebaho lug'atlardan biri sanaladi. Shu asarning "Indeks lug'ati" ham olimlarimiz tomonidan tayyorlangan. Mana shu lug'at bahona turkiy qabila-urug'larining tillari bir bobo tildan rivojlanib, taraqqiy etganligi haqidagi ilmiy-nazariy qarash dalillar asosida tasdiqlandiki, bu hol ona tilimizning o'ziga xos nufuzini va tarixiy taraqqiyotini dalillar vositasida tasdiqlab turibdi. [Ernazarov T.J, Karimjonova V.A, Ernazarova M.S, Mahmadiyev Sh.S, Rixsiyeva K.G', 2012; 2]. O'zbek lahja hamda shevalarining xususiyatlarini yaxshi bilish, o'rganib olish o'qituvchilar uchun, ayniqsa, "Ona tili va adabiyoti" fanining o'qituvchilari uchun ham juda zarur hisoblanadi. Ushbu fan o'qituvchisi o'z o'quvchilarining nutqidagi, yozuvidagi shevaga oid xususiyatlarni yaxshi bilsa, o'quvchilarning yozma va og'zaki nutqiga xos bo'lgan kamchiliklarini to'g'ri topa oladi. Buning uchun shevalarimiz haqida mukammal ma'lumotga ega bo'lishi shart. Shundagina, o'quvchilarning adabiy til talablari va me'yorlarini o'zlashtirib borish darajasidagi savodxonligi oshib boradi. [Ernazarov T.J, Karimjonova V.A, Ernazarova M.S, Mahmadiyev Sh.S, Rixsiyeva K.G', 2012; 13].

O'zbek tili shevashunosligi V.V.Reshetov, F.Abdullayev, Sh.Shoabdurahmonov, A.Aliyev, B.Jo'rayev, X.Doniyorov, Y.G'ulomov, Y.Jumanazarov, A.Ishayev, D.Abdurahmonov, N.Rajabov, Q.Muhammadjonov, N.Shoimova kabi dialektolog olimlar tomonidan chuqur o'rganib tahlil etilgan. Shuningdek prof. I. I.Zarubin, prof. K. K. Yudaxin; prof. G'ozzi Olim, prof. A. K. Borovkov, prof. V.V.Reshetovlar tomonidan shevalarimizni tasnif qilish bo'yicha olib borilgan ishlarni shevalarimizning mamlakatimiz va qo'shni davlatlarning hududlarida joylashishi jihatdan

tasnifiga misol bo‘la oladi (Enazarov T. J., Karimjonova V. A., Ernazarova M. S., Mahmadiyev SH. S, Rixsiyeva K. G‘, 2012; 23). Shevalar milliy tilimizning ilk poydevori bo‘lib, aynan shu shevalar orqali tilimiz boyiydi, hamda taraqqiy etadi. Professor E. D. Polivanov “O‘zbek dialektologiyasi va o‘zbek adabiy tili” nomli asarida (1933) o‘tgan asrning birinchi choragidayoq o‘zbek tili o‘zining ko‘p shevaliligi bilan boshqa turkiy tillardan ajralib turishi haqida ma’lumot bergan edi. Keyingi tadqiqotlar buni tasdiqlagan. Uning tub sabablari bir necha omillarga borib taqaladi. Professor V. V. Reshetov o‘zbek tili dialektal kartasining rang-barangligini uning etnogenezisidan qidirish lozimligini uqtiradi. Tariximizga oid adabiyotlar qadimda turkiy urug‘ va qabilalar miqdori ancha ko‘p bo‘lganligi to‘g‘risida xabar beradi. Ular ichida o‘zbek tili va uning shevalari shakllanishida qarluq, chigil, uyg‘ur, qipchoq, urug‘lari, qang‘li, qatag‘on urug‘ va qabilalari birlashmalarining tarixiy taraqqiyoti katta rol o‘ynagan. O‘zbek tilining dialektal o‘ziga xosligida qadimgi urug‘ va qabilalarning turli territoriyada alohida-alohida mustaqil yashashlari muhim ahamiyat kasb etgan. Ijtimoiy hayot tarzi, territorial mustaqilligi, kasb-kori, etnografik xususiyatlari ularning tilida aks etmay qolmas edi va bu holat tilda umumiylikni saqlab qolgan holda, boshqa territoriyada yashovchi o‘zbeklar tilidan ko‘pgina xususiyatlari bilan farqlanishiga olib kelgan, bu esa shevaning o‘ziga xosligini keltirib chiqargan. Sheva va dialekt xususiyatlarining tobora qat‘iylashishiga aholi punktlarining bir - biridan uzoq joylashishi muhim rol o‘ynagan. Shu tariqa bir til ichida turli shevalarning shakllanishiga olib kelgan.[Ashirboyev S, 2016; 7-8]. O‘zbek tiliga so‘zlar, asosan, uch manbadan: arab, fors-tojik, rus tilidan o‘zlashgan, lekin kamroq bo‘lsa-da boshqa tillardan ham o‘zlashgan so‘zlar bor. Bu so‘zlar shevalarda fonetik qulaylashtirilgan holda iste’molda bo‘ladi. Ilmiy-tadqiqot ishlari natijalari shuni ko‘rsatadiki, shevalar lug‘at tarkibida forscha-tojikcha so‘zlar ko‘proq, arabcha so‘zlar unga nisbatan kamroq, ruscha-baynalmilal so‘zlar juda kam uchraydi, bu esa bunday so‘zlarning o‘zlashish jarayoni ko‘proq vaqt talab qilishi bilan bog‘liq ekanini ko‘rsatadi.[Ashirboyev S, 2016; 52]. O‘zbek tilining shevalari orasida Buxoro shevasi markaziy (qarluq) lahjaga mansub bo‘lib, o‘ziga xos fonetik, morfologik va leksik xususiyatlari bilan ajralib turadi. Bu sheva asosan Buxoro viloyati va uning atrofidagi tumanlarda tarqalgan bo‘lib, tarixan o‘zbek adabiy tilining shakllanishida muhim rol o‘ynagan. Buxoro shevasining fonetik tizimi adabiy til bilan umumiylikka ega bo‘lsa-da, ayrim o‘ziga xosliklar bilan ajralib turadi. Eng muhim fonetik xususiyatlar quyidagilardan iborat:

1. Unli tovushlar tizimida o‘zgarishlar:

- Adabiy tilidagi “a” unlisining ba’zi so‘zlarda “o” tovushiga aylanishi:

adabiy til: baliq → shevada: boliq

adabiy til: sariq → shevada: soriq

- Ayrim hollarda “e” unlisi o‘rnida “i” yoki “a” tovushi ishlatiladi:

kelin → kilan, berdi → bardi

2. Undosh tovushlar tizimida o‘zgarishlar:

- “k” tovushining yumshoq talaffuzi yoki “x” tovushiga almashishi:

qilish → xiliş, ketdi → xetdi

- “g” tovushining yo‘qolishi yoki kuchsizlanishi:

bog‘ → bo‘, tog‘ → to‘

Quyidagi tahlilimizda Buxoro viloyati Vobkent tumanida yashovchi aholining kiyim – kechak hamda qarindosh urug‘ nomlariga tegishli bo‘lgan shevadagi so‘zlar va ularning adabiy tildagi varianti o‘rtasidagi farqlar ko‘rib chiqiladi.

Qarindosh – urug‘ nomlarining shevadagi va adabiy tildagi shakllari

T/R	Shevadagi so‘z	Adabiy tildagi shakli	So‘z turkumi
1.	Dada	Ota	Ot
2.	Nana	Ona	Ot
3.	Bobo	Buva	Ot

4.	Bibi	Buvi	Ot
5.	Apa	Opa	Ot
6.	Tog'o	Tog'a	Ot
7.	Yangamlo	Kelinoyi	Ot
8.	Yaznamlo	Pochcha	Ot
9.	Xesh	Qarindosh	Ot
10.	Apamlo	Qayinopa	Ot
11.	Ukamlo	Qayinuka	Ot
12.	Bacha	Bola	Ot

Kiyim – kechak nomlari

T/R	Shevadagi so'z	Adabiy tildagi shakli	So'z turkumi
1.	Kuylak	Ko'ylak	Ot
2.	Patinka	Poyabzal	Ot
3.	Poycha	Lozim	Ot
4.	Ichekcha	Nimcha	Ot
5.	Bruk	Shim	Ot
6.	Chapon	Chopon	Ot
7.	Kastum	Kastyum	Ot
8.	Jilop	Paypoq	Ot
9.	Daspo'shak	Qo'lpog	Ot
10.	Kalapo'sh	Do'ppi	Ot
11.	Ro'pokcha	Belbog'	Ot

Buxoro viloyatida tojik tilida so'zlashuvchilar ko'p bo'lganligi sababli sheva ham fors-tojik tili ta'sirida shakllangan. Hudud shevasida tojik tilining ta'sirini tahlil qilingan ishlarimizning kiyim-kechak nomlarida ko'p uchratishimiz mumkin. Vobkent shevasiga oid jihatlarni o'rganish orqali biz bu tilning shakllanishi hamda tillar bilan aloqadorligini tahlil qilish imkonini yaratamiz. Buxoro shevasi xalq og'zaki nutqida keng qo'llanadi va madaniy merosning bir qismi sifatida ahamiyatlidir. Taqqoslashlar asosida shuni aytish mumkinki, adabiy til va sheva o'rtasidagi farqlar nafaqat tilshunoslik, balki madaniyatshunoslik nuqtai nazaridan ham tadqiq etilishi lozim. Shevalarni o'rganish o'zbek tili taraqqiyotini chuqurroq anglash imkonini beradi.

Foydalanilgan adabiyotlar:

1. Ernazarov T.J, Karimjonova V.A, Ernazarova M.S, Mahmadiyev Sh.S, Rixsiyeva K.G' O'zbek dialektologiyasi (o'quv qo'llanma) Toshkent Universitet 2012
2. Ashirboyev, S. O'zbek dialektologiyasi. –Toshkent: Navro'z, 2016
3. Тўйчибоев Б., Ҳасанов Б. Ўзбек диалектологияси. –Т.: Абдулла Қодирий номидаги халқ мероси нашриёти, 2004.

O'ZBEK TILIDAGI ADABIY MATNLARDA SENTIMENT ANALIZ IMKONIYATLARI

Rahimov Xasanboy

Andijon davlat universiteti 3-bosqich tayanchdoktoranti
Namangan davlat chet tillar instituti stajyor o'qituvchisi

Annotatsiya: Mazkur tadqiqotda o'zbek tilidagi adabiy matnlar, xususan she'riy asarlarda sentiment analiz qilish imkoniyatlari o'rganiladi. Tadqiqotda o'zbek adabiyotiga xos emotsional ifoda vositalari tahlil qilinib, sentiment analiz uchun leksik va kontekstual yondashuvlar taqdim etiladi. Shuningdek, adabiy matnlar sentimentini aniqlash uchun dastlabki korpus yaratish va sentiment lug'atini tuzish usullari taklif qilinadi. Tadqiqot natijalari, o'zbek adabiyotidagi emotsional tendensiyalarni zamonaviy sun'iy intellekt asosida tahlil qilish orqali adabiy matnlarni o'quvchilar yoshidan kelib chiqib tavsiya qilish imkoniyatlarini ko'rsatadi.

Kalit so'zlar: sentiment analiz, korpus, leksikon, NLP, kontekstual yondashuv.

Annotation: This study explores the possibilities of conducting sentiment analysis in Uzbek literary texts, particularly in poetic works. The research analyzes the emotional expressive means characteristic of Uzbek literature and presents lexical and contextual approaches for sentiment analysis. In addition, the paper proposes methods for creating an initial corpus and compiling a sentiment lexicon for determining sentiment in literary texts. The results of the study demonstrate the possibilities of analyzing emotional tendencies in Uzbek literature using modern artificial intelligence, and suggest recommending literary texts according to the reader's age group.

Keywords: sentiment analysis, corpus, lexicon, NLP, contextual approach.

Аннотация: В данном исследовании рассматриваются возможности проведения сентимент-анализа в узбекских литературных текстах, в частности в поэтических произведениях. В работе анализируются эмоционально-выразительные средства, характерные для узбекской литературы, и предлагаются лексический и контекстуальный подходы к сентимент-анализу. Также предлагаются методы создания начального корпуса и составления сентимент-лексикона для определения сентимента в литературных текстах. Результаты исследования демонстрируют возможности анализа эмоциональных тенденций в узбекской литературе с использованием современных систем искусственного интеллекта и дают рекомендации по подбору литературных произведений в зависимости от возраста читателей.

Ключевые слова: сентимент-анализ, корпус, лексикон, NLP, контекстуальный подход.

Bugungi har tomonlama rivojlangan davrda badiiy asarlar ham alohida o'rganishga va tahlil qilishga muhtojdir. Ularda qo'llangan so'zlardagi ijobiy yoki salbiy ma'no ustuvorligiga qarab asar ruhiyati, qaysi yoshdagi o'quvchiga xosligi va undagi so'z va iboralar qo'llanilishiga qarab asarning murakkablik darajasi avtomatik tarzda aniqlanishi mumkin. Buni amalga oshirish uchun sentiment analizdan foydalaniladi. Sentiment analiz — matnlarda ifodalangan kayfiyat, hissiyot va emotsional munosabatlarni avtomatik aniqlash va tahlil qilish usuli hisoblanadi. U ko'proq ijtimoiy tarmoqlar, yangiliklar va foydalanuvchi fikrlarini tahlil qilishda qo'llaniladi [Pang, B., & Lee, L. 2008]. Shunga qaramay adabiy matnlar ustida ham sentiment analiz o'tkizish mumkin va bu jamiyat uchun yuqorida aytib o'tilganidek foydalidir. Adabiy matn — estetik va badiiy vositalar yordamida ifodalangan, ma'lum badiiy maqsadga yo'naltirilgan lingvistik birliklar majmui bo'lib, unda hissiyot va ruhiy kechinmalar muhim o'rin egallaydi [Qo'chqorov, O. 2019]. She'riyat esa bu xususiyatlar eng yuqori darajada jamlangan janr hisoblanadi. Adabiy matnarni sentiment analiz qilishning turli yo'llari mavjud. Hozirgi zamon kompyuter lingvistikasi va NLP

(Natural Language Processing) sohalarida sentiment analizning ikki asosiy usuli mavjud: 1. Leksik asoslangan yondashuv — matndagi emotsional soʻz va iboralarni tayyor sentiment lugʻat asosida aniqlash [Liu, B. 2012]. 2. Kontekstual yondashuv — soʻz va iboralar maʼnosini kontekstga qarab avtomatik belgilash va emotsional baholash [Cambria, E., & White, B. 2014]. Bu ikki usulni amalga oshirish uchun tadqiqotchiga sentiment lugʻati va korpus kerak boʻladi. Korpus — bu til birliklarini tizimli ravishda yigʻish, tasniflash va tahlil qilish imkonini beruvchi elektron matnlar toʻplamidir. U lingvistik tadqiqotlar uchun asosiy manba boʻlib xizmat qiladi [Abdurahmonova, N. Z., 2023. 346]. Sentiment lugʻati esa — tilning ijobiy, salbiy va neytral hissiy soʻzlarini oʻz ichiga oluvchi maxsus lugʻat boʻlib, matnning emotsional tonini belgilashda asosiy vosita hisoblanadi [Tursunov, S., & Karimov, D., 2021]. Hozir Abdulla O.ning “Yillar armoni” [Oripov, A., 2001] deb nomlangan kichik sheʼrini ham leksik ham kontekstual uslubda sentiment analiz qilib koʻramiz. Ushbu sheʼrda inson hayotining oʻtkinchiligi, qilinmay qolgan ezgu ishlar va umr davomidagi afsuslar ifodalanadi. Sheʼr leksikasida “yaxshi gap”, “rahmat”, “uzr”, “taskin” kabi ijobiy ohangdagi birliklar bilan bir qatorda, “ranjib”, “qusurlar”, “armon”, “alamlar” kabi salbiy kayfiyatdagi soʻzlar mavjud. Leksik tahlil natijasida, salbiy va ogʻir tuygʻular ustunligi aniqlangan boʻlsa-da, yakunda taskin va afsus motivlari orqali ruhiy yengillik berilishi kuzatiladi. Bu usulda har bir soʻzga sentiment skori (bahosi) beriladi. Masalan: ijobiy soʻzlar: +1, salbiy soʻzlar: -1, neytral soʻzlar: 0.

$$\text{Formulasi: } s = \frac{\sum_{i=1}^n s(w_i)}{n}$$

Bu yerda: s — matnning umumiy sentiment skori; $s(w_i)$ — soʻzning skori; n — umumiy soʻzlar soni.

Agar: $s > 0$ boʻlsa — ijobiy
 $s < 0$ boʻlsa — salbiy
 $s = 0$ boʻlsa — neytral

Kontekstual sentiment tahlilda esa, “Yillar armoni”, “Armon koʻrmadim sizdan buyukroq” kabi metaforalar inson ruhiyati va oʻtgan umr haqidagi falsafiy iztirobni ifodalaydi. Sheʼr mantiqiy jihatdan uch qismga boʻlinadi: kirish qismi falsafiy achinish, asosiy qismi iztirob va armon, yakun esa taskin va ruhiy yengillik bilan yakunlanadi. Bu usulda soʻzlar kontekstdagi vazifasiga qarab baholanadi. Masalan: “Armon” yolgʻiz boʻlsa: -1; “Taskin” “armon” bilan bir gapda boʻlsa: +0.5.

Bu yerda koeffitsientlar kontekstga mos ravishda belgilanadi.

$$\text{Formulasi: } S_c = \frac{\sum_{j=1}^m S_c(p_j)}{m}$$

Bu yerda: S_c — kontekstual sentiment skori

$S_c(p_j)$ — j -chi kontekst boʻlagi uchun skori

m — umumiy kontekst boʻlaklari soni

Quyida TextBlob va NLTK yordamida soddalashtirilgan sentiment analiz kodi:

```
from textblob import TextBlob

# Sheʼr matni
text = """Kimqadir yaxshi gap aytdik, birodar,
Kimdir bir umrga ranjib ham ketdi.
Kun kecha yosh derdik oʻzni bir qadar,
Bugun-chi baloqʻat yoshi ham yetdi."""

# TextBlob yordamida sentiment baholash
blob = TextBlob(text)
sentiment = blob.sentiment.polarity

# Natija
if sentiment > 0:
    print("Ijobiy kayfiyat")
elif sentiment < 0:
    print("Salbiy kayfiyat")
else:
    print("Neytral")

print("Sentiment skori:", sentiment)
```

Yuqoridagi kod orqali she'r matni tahlil qilinib, -1 dan +1 gacha **polarity** skori, ya'ni sentiment analiz natijasi chiqadi.

Abdulla O. she'ri sentiment analizida salbiy-falsafiy kayfiyat bilan boshlanib, ijobiy taskin bilan tugallanadigan asar sifatida baholanadi. Ushbu yondashuv o'zbek adabiyotidagi emotsional tendensiyalarni tahlil qilish va sun'iy intellekt uchun milliy sentiment lug'atlarini yaratishda muhim asos bo'la oladi. O'zbek tilida yaratilgan adabiy matnlarni ko'p miqdorda yeg'ib, ular asosida korpus va sentiment lug'at tashkil qilish orqali mazkur matnlar ustida bajarilishi mumkin bo'lgan sentiment analiz imkoniyatlaridan kengroq miqyosda foydalanish imkoniyatini qo'lga kiritiladi.

Foydalanilgan adabiyotlar.

1. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/15000000011>
2. Qo'chqorov, O. (2019). Adabiy til va uslub masalalari. — Toshkent: Fan va texnologiya nashriyoti.
3. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1–167.
4. Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational intelligence magazine*, 9(2), 48–57.
5. Abdurahmonova, N. Z. (2023). *Korpus lingvistikasi*. Toshkent: O'zbekiston Milliy universiteti. 346 b.
6. Tursunov, S., & Karimov, D. (2021). O'zbek tilida sentiment lug'ati yaratish tajribalari. *Filologiya masalalari*, №4, 45–52.
7. Abdulla Oripov. Yillar armoni // Abdulla Oripov. Saylanma asarlar. — Toshkent: G'afur G'ulom nomidagi Adabiyot va san'at nashriyoti, 2001. — 1-jild. — B. 314-315.
8. Rahimov H. Behavioral lexemes and their importance in sentimental analysis // "Western European journal of linguistics and education". Volume 2, Issue 11 November 2024. Germany P. 152-157
9. Raximov X. Sentiment tahlil uchun o'zbekcha matnlarni tasniflash me'yorlari // "Published of Science and Education" 01.01.2025 P.21-26
10. Raximov X. Tabiiy tilni qayta ishlash(NLP)da korpus texnologiyalarini o'rni . // "Zamonaviy dunyoda innovatsion tadqiqotlar: Nazariya va amaliyot" nomli ilmiy, masofaviy onlayn konferensiya. 2025 yil 1(56)-son B.16-21
11. Rahimov H. Analysis of global experience in text analysis //Texas Journal of Philology, Culture and History. Januar 2025. P.109-113

O'ZBEK TILIDAGI OTLARNING CONCEPTICON ASOSIDA KONSEPTUAL XARITASINI YARATISH MASALALARI

Tirkashev Jasurjon Uktam o'g'li

O'zbekiston Milliy universiteti mustaqil tadqiqotchisi

E-mail: jasurbey77@gmail.com

Annotatsiya: Mazkur ilmiy ishda o'zbek tilidagi 1000 ta ot so'zning semantik tahlili Concepticon konseptual bazasi asosida amalga oshirilgan. Tadqiqotda otlar Concepticondagi mavjud kategoriyalarga muvofiq tartiblanib, ularning semantik guruhlari aniqlangan. So'zlar ko'p kategoriyali yoki yagona kategoriyali bo'lishi mumkinligi tahlil qilinib, ularning semantik tuzilmasi grafik ko'rinishda tasvirlangan. Ishda shuningdek, o'zbek otlarining madaniy, ijtimoiy va psixologik konseptlar bilan bog'liqligi, ularning boshqa tillardagi konseptual tizimlar bilan o'xshash va farqli jihatlari ham muhokama qilingan. Tadqiqot natijalari o'zbek tilshunosligida konseptual tahlilga asoslangan yangi yondashuvlar uchun nazariy va amaliy asos yaratadi.

Kalit so'zlar: otlar, konseptual xarita, concepticon, semantik tahlil, konsept, leksik-semantik guruh, kognitiv lingvistika, ma'no tarmoqlari, semantik tasnif

Abstract: This research presents a semantic analysis of 1,000 Uzbek nouns using the Concepticon conceptual database. The nouns are categorized according to the classifications provided by Concepticon, and their semantic groupings are identified. The analysis reveals whether words are polycategorical or belong to a single category, and their semantic structures are visualized through conceptual mapping. The study also discusses the cultural, social, and psychological associations of Uzbek nouns and compares them with conceptual systems in other languages. The findings provide both theoretical and practical foundations for introducing concept-based approaches in Uzbek linguistics.

Keywords: nouns, conceptual map, concepticon, semantic analysis, concept, lexical-semantic group, cognitive linguistics, meaning networks, semantic classification

Аннотация: В данном научном исследовании проведён семантический анализ 1000 существительных узбекского языка на основе концептуальной базы данных Concepticon. Существительные были классифицированы в соответствии с категориями, представленными в Concepticon, и определены их семантические группы. Анализ выявил, являются ли слова многокатегориальными или принадлежат к одной категории, а также представлены графические схемы их семантической структуры. В исследовании также рассматриваются культурные, социальные и психологические аспекты, связанные с узбекскими существительными, и проводится сопоставление с концептуальными системами других языков. Результаты исследования создают теоретическую и практическую основу для внедрения концептуального подхода в узбекском языкознании.

Ключевые слова: существительные, концептуальная карта, concepticon, семантический анализ, концепт, лексико-семантическая группа, когнитивная лингвистика, сети значений, семантическая классификация

Kirish

O'zbek tilining lug'at tarkibini semantik jihatdan chuqur o'rganish va tasniflash zamonaviy tilshunoslikning muhim yo'nalishlaridan biridir. So'zlarning ma'nolarini formal shaklda ifodalash va ularni turkumlash uchun turli usullar mavjud. Masalan, ma'nodoshlik tarmoqlarini yaratishga asoslangan WordNet tizimi keng qo'llaniladi. Unda so'zlar sinonimik guruhlar shaklida, ya'ni konseptual ma'no birliklari tarzida tartiblangan [uzresearchers.com]. Shu bilan birga so'zlarni ma'nolariga ko'ra tasniflash va ularni turli konseptlar xaritasiga joylashtirishning boshqa usuli ham mavjud. Xususan, xalqaro lingvistik platformalarda Concepticon deb nomlanuvchi maxsus

ma'lumotlar bazasi taklif qilingan bo'lib, u turli tadqiqotlardagi konseptlar ro'yxatini yagona standartlashtirilgan konseptlar to'plamiga bog'lashga xizmat qiladi [concepticon.clld.org]. Concepticon asosan "qarindosh konseptlar" majmuasini hosil qiladi va tillarni qiyosiy o'rganishda umumiy tushunchalarga tez murojaat qilish imkonini beradi [concepticon.clld.org].

O'zbek tilshunosligida hozircha konseptual xaritalar va konseptlar ro'yxati bilan ishlaydigan bunday resurslar cheklangan. Garchi so'nggi yillarda kompyuter lingvistikasi va korpus lingvistikasi jadal rivojlanib, milliy korpuslar yaratilgan bo'lsa-da, leksik birliklarni konseptual tasniflash masalasi yetarlicha yoritilmagan. Shu nuqtai nazardan, N. Abdurahmonova tomonidan O'zbek tilining elektron korpusi va uning kompyuter modellari haqida yozilgan tadqiqot ishida korpusdan olingan lug'aviy birliklarni turli ma'lumotlar bazalariga bog'lash masalasi ham ko'tarilgan [academia.edu]. Shuningdek, o'zbek tilida axborot-qidiruv tizimlari uchun lingvistik ontologiya va tezaurus yaratish bo'yicha dastlabki izlanishlar amalga oshirilgan bo'lib, unda konseptlar va ularning o'zaro munosabatlarini aks ettiruvchi ontologik model taklif qilingan. Bu kabi ilmiy ishlar o'zbek tilining konseptual asoslangan elektron lug'at va bazalarini yaratishga bo'lgan ehtiyoj ortib borayotganini ko'rsatadi.

Ushbu maqolada o'zbek tilidagi eng ko'p ishlatiladigan 1000 ta otni konseptual kategorialarga bo'lish va ularning "konseptlar xaritasi" ni tuzish masalasi ko'rib chiqiladi. Bunda xalqaro Concepticon konseptlar bazasi resursidan foydalanilgan holda o'zbek tilidagi otlar uchun umumlingvistik konseptual toifalar aniqlanadi. Otlarni ma'no jihatdan tasniflab, uning konseptual tuzilishini chiqarish, turli semantik kategoriyalar kesimida eng muhim konseptlarni ajratish va keyingi tadqiqotlar uchun asos yaratish maqsad qilingan. Shuningdek, ushbu ish davomida kuzatilgan muammolar tahlil qilinib, kelgusida o'zbek tili uchun konseptual asoslangan lingvistik resurslar yaratish istiqbollari muhokama qilinadi.

Tadqiqot materiallari va metodlari

Tadqiqot materiali sifatida o'zbek tilidagi eng ko'p qo'llaniladigan 1000 ta ot tanlab olindi. Ushbu so'zlar ro'yxati milliy matnlar korpusidan avtomatik ravishda ajratib olingan bo'lib, tez-tez ishlatiladigan otlarni qamrab oladi. Xususan, o'zbek milliy korpusidagi turli uslub va janrlarga mansub matnlardan morfologik analiz yordamida otlar ajratib chiqildi va ularning lemmasi bo'yicha saralandi. O'zbek tilining anglunativ tuzilishi sababli korpusdan so'zlarni ajratishda maxsus morfologik analizator va lemmatizatoridan foydalanildi. Bu borada yaratilgan dasturiy vositalar so'zning asosiy shaklini aniqlash va uni lug'aviy bazadagi normal shaklga keltirishga yordam berdi. Natijada 1000 ta tez-tez uchraydigan ot ro'yxatga olindi. Masalan, "*kitoblar*" emas "*kitob*", "*daryoni*" emas "*daryo*" va hokazo.

So'zlarni konseptlarga moslashtirish jarayoni ikki bosqichda amalga oshirildi: 1) Avtomatik bosqich – bunda har bir o'zbekcha otning ingliz tilidagi eng yaqin tarjimai yoki ekvivalenti aniqlanib, Concepticon ma'lumotlar bazasidan shu so'zga mos konsept topish amalga oshirildi; 2) Qo'lda tekshirish bosqichi – dastlabki avtomatik moslash natijalar ko'rib chiqilib, noto'g'ri yoki noaniq moslashlar tuzatildi. Avtomatik moslashtirish uchun asosan ikki xil manba ishlatildi. O'zbekcha-inglizcha ikki tilli lug'atlar bazasi va zarur hollarda Google translate kabi mashina tarjimai vositasi yordamida so'zning inglizcha muqobili olindi. So'ngra o'sha inglizcha so'z Concepticonda mavjud konseptlar ro'yxatidan qidirildi. Concepticon bazasi hozirgi vaqtda 4000 tacha standart konseptual birlikni o'z ichiga oladi [concepticon.clld.org]. Har bir konsept biror umumiy tushunchani ifodalaydi va turli tillardagi muqobil so'zlar u bilan bog'langan. Shunday qilib o'zbek tilidagi har bir ot uchun uning ifodalayotgan konseptiga mos keluvchi Concepticon konsept tarmoq aniqlab borildi.

Mazkur jarayonda ayrim so'zlar bir nechta konseptga mos kelishi yoki umuman mos konsept topilmasligi holatlari kuzatildi. Bunday vaziyatda qo'shimcha qo'lda tahlil qilindi. So'zning turli ma'nolari ajratilib, qo'llanilish chastotasiga qarab, ma'nosi tanlandi va shu ma'no uchun mos konsept belgilandi. Masalan, o'zbek tilidagi "*oy*" so'zi ikki ma'noga ega: *osmondagi yorug' planeta* "*moon*" va *yilning 12-dan biri*, *oy* "*month*". Korpusdagi qo'llanishlar tahlil qilinib,

ko'proq qaysi ma'noda ishlatilishi aniqlanadi. Natijada, "oy" so'zi "month" konseptiga moslashtirildi, chunki kundalik matnlarda vaqt birligi ma'nosi ko'proq uchradi.

Konseptual toifalarga ajratish. Olingan konseptlar ro'yxati keyingi bosqichda ma'no doirasiga ko'ra kategoriyalarga ajratildi. Bu kategoriyalar semantik maydonlarni ifodalaydi. Tadqiqot ehtiyojiga mos 10 ta keng kategoriyani belgiladik:

- 1) *Inson va jamiyat* (odam, qarindoshlik, ijtimoiy holat, millat, kasb-hunar va boshqalar),
- 2) *Tabiat* (hayvonot va o'simliklar dunyosi, tabiiy hodisalar),
- 3) *Jismoniy obyektlar* (sun'iy narsalar, asbob-anjom, uy-ro'zg'or buyumlari, transport vositalari va hk),
- 4) *Joy va makon* (joy nomlari, makon tushunchalari: uy, shahar, davlat, olam...),
- 5) *Vaqt* (kun, yil, vaqt birliklari, davrlar),
- 6) *Tana a'zolari va sog'liq* (inson tanasi, kasalliklar),
- 7) *Faoliyat va jarayon* (masalan, "terim", "o'yin", "uyqu"),
- 8) *Mulk va iqtisod* (mol-mulk, pul, iqtisodiy tushunchalar),
- 9) *Mavhum tushunchalar* (mavhum kategoriyalar: hayot, baxt, haqiqat, muammo, fikr va sh.k.),

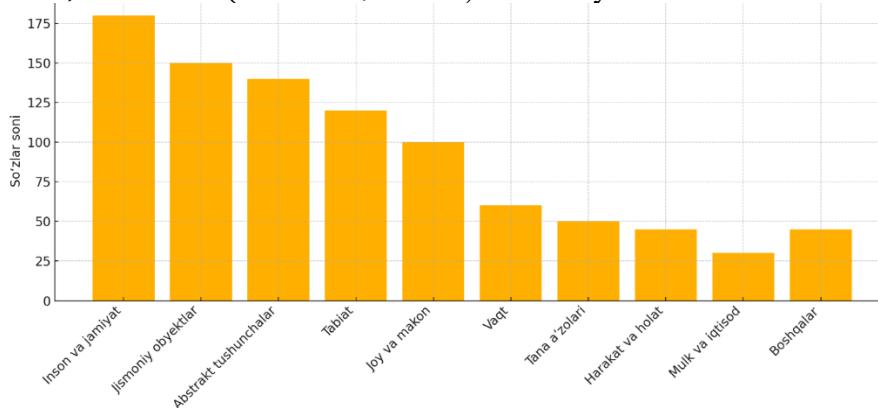
10) *Boshqalar* (yuqoridagilarga kirmagan qolgan umumiy otlar: masalan, til nomlari, ranglar, va b.). Kategoriyalar sonini nisbatan kamroq qilish orqali har bir guruhga ko'proq so'zlarni jamlash va umumiy tendensiyani ko'rish maqsad qilingan. So'z-konsept juftliklari shu keng toifalarga tegishlilik bo'yicha ajratilib chiqildi.

Natijalar

Konseptlarga moslashish qamrovi: 1000 ta eng ko'p qo'llanuvchi ot 918 tasi (92%) Concepticondagi mavjud konseptlarga muvaffaqiyatli mos keldi. Boshqa 82 ta so'zga to'g'ridan-to'g'ri konsept topilmadi. Mos kelmagan so'zlar asosan maxsus nomlar, masalan, "Toshkent", "Alloh" kabi o'ziga xos nomlar yoki madaniy tushunchalar, masalan, "mahalla", "hashar" singari faqat o'zbek xalqiga xos atamalar bo'lib chiqdi. Ushbu so'zlar Concepticon doirasida umumiy konsept sifatida qaralmagan, chunki ular global konseptlar ro'yxatlarida uchramaydi. Shunday qilib tadqiq qilingan korpusdagi eng faol otlarning katta qismi (92 foizi) umumjahon konseptual tushunchalar doirasida ifodalanishi mumkinligi aniqlandi. Bu bir tomondan o'zbek tilining eng asosiy lug'ati ko'plab boshqa tillardagi kabi umumiy tushunchalarni qamrab olishini ko'rsatadi. Boshqa tomondan, 8% atrofidagi qismi faqat o'zbek tiliga yoki mintaqaga xos leksik birliklar ekanini ham ta'kidlab o'tish lozim.

Semantik kategoriyalar bo'yicha taqsimot: Yuqorida belgilangan 10 ta konseptual kategoriya kesimida so'zlar soni quyidagicha taqsimlandi (1-rasmga qarang). Eng yirik guruh "Inson va jamiyat" bo'lib, unga 180 ta (taxminan 18%) ot kirdi. Bu guruhga odam, qarindoshlar (*ota, ona, aka, uka, opa, singil, farzand* va hokazo), shaxslar (*erkak, ayol, bola, do'st, dushman*), kasb va unvonlar (*o'qituvchi, shifokor, talaba, prezident*), jamiyat va millatga oid tushunchalar (*xalq, millat, davlat, hukumat*) kabi so'zlar jamlandi. Ikkinchi yirik kategoriya "Jismoniy obyektlar" bo'lib, 150 ta (15%) so'zni o'z ichiga oldi. Bu guruhga uy-ro'zg'or buyumlari (*stol, stul, idish*), transport vositalari (*mashina, poyezd*), texnika va texnologiya bilan bog'liq otlar (*telefon, kompyuter*), kiyim-kechak va boshqalar kiradi. Keyingi o'rinda "Mavhum tushunchalar" kategoriyasi 140 ta (14%) bilan kelmoqda. Bular inson tafakkuridagi mavhum holatlar (*hayot, o'lim, baxt, muammo, fikr, haqiqat, sevgi* kabi). "Tabiat" kategoriyasiga 120 ta so'z (12%) kirdi. Hayvonlar (*it, mushuk, ot, sigir, qush*), o'simliklar (*daraxt, gul*), tabiat hodisalari (*yer, suv, olov, tog', quyosh, oy, yulduz, kun* va hokazo). "Joy va makon" bo'yicha 100 ta so'z (10%) aniqlandi. Turli turar-joy va geografik atamalar (*uy, ko'cha, shahar, mamlakat, dunyo* va hk). "Vaqt" kategoriyasiga 60 ta so'z (6%) tegishli bo'lib, yil fasllari, hafta kunlari, vaqt o'lchovlari va davrlarni ifodalaydi (*yil, oy, hafta, kun, bahor, yoz* va boshqalar). "Tana a'zolari va sog'liq" guruhida 50 ta so'z (5%) bor. Inson tanasi qismlari (*bosh, qo'l, ko'z, yurak, qon*), kasallik va tibbiyotga oid so'zlar (*isitma, dori* va hk). "Faoliyat va jarayon" kategoriyasiga 45 ta (4.5%) ot kirdi (*o'yin, jang, sayohat, uyqu, musoboqa* kabi jarayon nomlari). "Mulk va iqtisod" bo'yicha 30

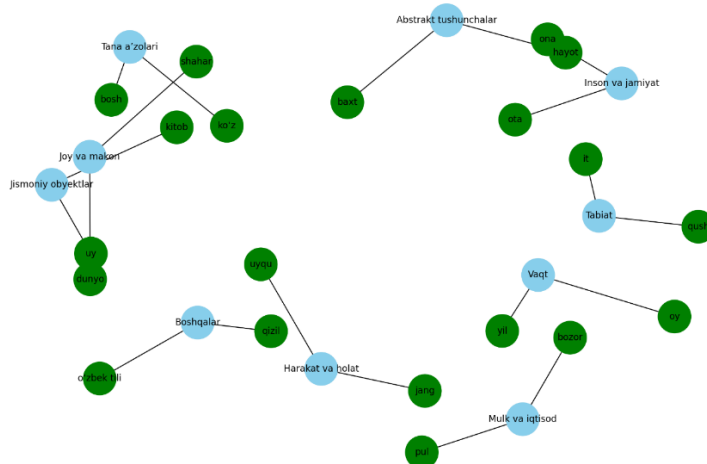
ta (3%) soʻz chiqdi. Mol-mulk va iqtisodiy tushunchalar (*pul, boylik, bozor, narx, ish* va hk). Nihoyat, “Boshqa” turkumiga 45 ta (4.5%) turli xil nomlar kiritildi. Bular yuqoridagilarga kirmagan, masalan, til nomlari (*oʻzbek tili, rus tili*) va baʼzi yondosh tushunchalardir.



1-rasm. Oʻzbek tilidagi 1000 ta eng koʻp ishlatiladigan otning semantik kategoriyalar boʻyicha taqsimoti. Har bir ustun muayyan konseptual toifaga kiruvchi soʻzlarning sonini koʻrsatadi.

Yuqoridagi diagrammadan koʻrinib turibdiki, inson va jamiyat sohasi oʻzbek tilidagi otlarda eng katta ulushga ega. Bu tabiiy hol, chunki eng koʻp qoʻllanadigan soʻzlarning orasida insonning oʻzi, uning atrofidagi odamlar, jamiyat va davlat haqidagi tushunchalar doimo markaziy oʻrin tutadi. Keyingi yirik kategoriyalar, sunʼiy obyektlar va abstrakt tushunchalar, mos ravishda ikkinchi va uchinchi oʻrinda. Bu shuni anglatadiki, odamlar kundalik muloqot va matnlarda koʻp hollarda jismoniy narsalar nomini va mavhum tushunchalarni tilga oladilar. Tabiat va joy-makon kategoriyalari ahamiyatli ulushga ega, har biri 10% atrofida. Bu esa atrof-muhit, geografiya va joy tushunchalari ham tez-tez muhokama qilinishini koʻrsatadi. Vaqt tushunchalari nisbatan kamroq 10%, ammo baribir muhim qism. Tana aʼzolari va faoliyat otlari ozroq ulushga ega, chunki bu maʼnolar koʻpincha feʼl yoki sifat bilan ifodalanadi, ot tarzida kamroq qoʻllanadi. Mulk-iqtisod soʻzlari esa eng kichik guruhlardan, bu maxsus matnlarda koʻproq qoʻllanadi, masalan, iqtisodiy matnlarda.

Soʻz-konsept tarmogʻi: 2-rasmda namunaviy ravishda 20 ta ot va ularning bogʻlangan konseptual kategoriyalari tarmoq grafi koʻrinishida keltirilgan. Unda har bir yashil doira aniq ot, moviy toʻgʻri toʻrtburchak esa u tegishli boʻlgan konseptual toifani ifodalaydi. Chiziqlar soʻzning qaysi kategoriya bilan aloqador ekanini koʻrsatadi. Masalan, “ona” soʻzi “Inson (oil)” tuguniga ulangan, “qoʻl” soʻzi “Tana” tuguniga, “kitob” soʻzi “Obyekt” tuguniga, “baxt” soʻzi “Abstrakt” tuguniga ulangan va hokazo. Bu koʻrinishdan koʻrinadiki, tanlangan soʻzlarning orasida turli kategoriyalar mavjud va ular tarmoq shaklida oʻzbek tilining konseptual xaritasini hosil qiladi.



2-rasm. Ot soʻzlarning namunaviy konseptual tarmogʻi. Yashil doiralar – oʻzbek tilidagi soʻzlarning, moviy doiralar – ularga mos konseptual kategoriyalar.

Muhokama

Yuqoridagi natijalar o'zbek tilidagi otlar tarkibida inson va jamiyatga oid tushunchalar yetakchi o'rin tutishini ko'rsatdi. Bu kutilgan hol bo'lib, boshqa ko'plab tillarda ham eng faol otlar orasida *inson*, *oila*, *jamiyat* kabi konseptlar ustunlik qiladi. Masalan, rus yoki ingliz tillarini olsak ham, "*odam*", "*bola*", "*uy*", "*yil*", "*hayot*" kabi so'zlar eng ko'p ishlatiladigan birliklar sirasiga kiradi. Demak, bazaviy konseptlar katta darajada universalidir. Concepticon bazasi orqali o'zbek tilidagi so'zlarning 92% umumiy konseptlarga mos kelgani ham buning tasdig'idir.

Biroq har bir tilga xos bo'lgan leksik birliklar ham mavjudligi aniqlandi. Bu so'zlar konseptual xaritada "yopiq joylar" hosil qiladi, ya'ni ularning global konseptlarga to'g'ri keladigan ekvivalenti yo'q. Bunga misol tariqasida "*mahalla*" tushunchasini keltirish mumkin. Boshqa tillarda bevosita mos so'z bo'lmasa-da, "*community*" yoki "*neighborhood*" kabi kengroq konseptlar orqali ifodalanadi. Concepticonda esa "*mahalla*" degan alohida konsept yo'q, shu bois o'zbek tilidagi bu so'zni xaritada joylashtirish qiyin. Bu hol madaniy va ijtimoiy xoslikni aks ettiruvchi leksik birliklarni aniqlash va ularni alohida o'rganish zarurligini ko'rsatadi. Kelgusida Concepticon singari resurslarni boyitishda bunday xos konseptlarni ham kiritish masalasi turadi.

Mazkur tadqiqot davomida qator muammolar ham kuzatildi. Avvalo, ko'p ma'noli so'zlarni konseptlarga moslashtirishda noaniqliklar bo'ldi. Masalan, yuqorida keltirilgan "*oy*" so'zidan tashqari, "*bosh*" so'zi ham "*head*" – tana a'zosi va "*leader/main*" – rahbar ma'nolarini ifodalaydi. Shuningdek, "*tong*" – "*dawn*" va "*east*" ma'nolarida, "*til*" – "*tongue*" – organ va "*language*" – muloqot vositasi ma'nolarida ishlatilishi mumkin. Bu kabi hollarda korpusdagi eng ko'p ishlatilishiga qarab bitta ma'no tanlandi, lekin aslida bu so'zlar bir nechta konseptga bog'lanishi mumkin. Ko'p ma'nolilik muammosi konseptual xarita tuzishda alohida e'tibor talab etadi. Ehtimol, kelajakda har bir so'z uchun bir nechta konsept bog'lash va keyin ularning kontekstga qarab aniqlanish usullarini ishlab chiqish lozim bo'ladi.

Yana bir muammo – morfologik variantlar va so'z turkumlari o'rtasidagi chegaralarni hisobga olish. Bizning ishimiz faqat ot so'zlari bilan cheklangani uchun ayrim konseptlar chetda qoldi. Masalan, "*yaxshi*", "*katta*", "*bor*" kabi juda faol leksemalar sifat, modal so'z turkumlariga mansub bo'lgani sababli ro'yxatga olinmadi. Ammo aslida bular ham muayyan konseptlarni ifodalaydi. Demak, kelgusida boshqa so'z turkumlarini ham qamrab olgan holda konseptual xarita tuzish yanada to'liq manzara beradi. Shuningdek, otlarning yakka va ko'plik, egalik shakllari ham korpusda tez-tez uchraydi. Biz ularni lemmatizatsiya qildik, lekin real foydalanishda, masalan "*otam*" (*mening otam*) va "*ota*" (*umuman ota*) ma'noda farqlanishi mumkin. Bunday holatlar ayniqsa tarjimada muhim.

O'zbek tilining xususiyatlari konseptual xarita hosil qilishda qanday aks etdi? Birinchidan, o'zbek tili anglunativ bo'lgani tufayli birgina ot turli shakllarda yuzlab variantlarga ega bo'lishi mumkin. Bu muammoni lemmatizatsiya bilan hal etdik, lekin bu albatta dastlabki bosqichda qo'shimcha ish yukini berdi. Ikkinchidan, o'zbek tilida arabcha va forsha olinma otlar ko'p ishlatiladi, masalan, *kitob*, *maktab*, *vaqt*, *hayot* va *hokazo*. Ba'zi hollarda bunday so'zlarning sinonimlari ham bor, masalan, *hayot* – *tiriklik*, *vaqt* – *zamon* kabi. Bizning faol ro'yxatda odatda eng ommabop varianti qoldi.

Ushbu tadqiqot natijalari lingvistik resurslar yaratish nuqtai nazaridan muhim ahamiyatga ega. Birinchidan, 1000 ta eng faol otning konseptual xaritasini olish bilan biz lug'aviy ma'no tarmog'ining asosini ko'rdik. Buni yanada kengaytirib, boshqa so'z turkumlarini ham qo'shish orqali to'liq o'zbek tili konseptual tarmog'ini yaratish mumkin. Bu tarmoq WordNet singari sinonim tarmoqlaridan farqli o'laroq, tushuncha markazda bo'ladi va har bir konsept atrofida o'sha tilning turli leksemalarini birlashtiradi. Bunday yondashuv ko'ptillik tadqiqotlar uchun ayniqsa foydali, chunki Concepticon orqali dunyoning qirqdan ortiq tilidagi konseptlar bilan solishtirish imkoni paydo bo'ladi [concepticon.clld.org].

Kelgusida konseptga asoslangan lug'atlar va ma'no tezauruslarini yaratish o'zbek tili uchun dolzarbdir. Hozirda mavjud WordNet loyihasi sinonimiya va antonimiya bog'lanishlarini ko'rsatsa, konseptual lug'at har bir so'zning konseptual toifalarini va hatto ontologik

munosabatlarini berishi mumkin. Masalan, *ona* konsepti *ota-ona* konseptining qismi ekani kabi munosabatlar Concepticonda qisman bor, lekin ularni o'zbek tilida ham ifodalash va boyitish mumkin [concepticon.clld.org]. M. Abjalovanning tadqiqotida taklif etilgan lingvistik ontologiya aslida shunday tuzilma bo'lib, u ma'lum bir sohada, masalan, axborot izlashda foydali bo'lsa, uni umumtilshunoslikka ham tatbiq etish mumkin. Yana bir istiqbolli yo'nalish – mashina tarjimasi va sun'iy intellekt tizimlarida konseptual xaritadan foydalanish. Agar o'zbekcha matndagi so'zlar konseptlar bilan belgilanib borilsa, mashina semantik tushunish darajasiga erishishi osonlashadi. Masalan, ma'lum bir kontekstda “*olma*” so'zi “o'simlik” konsepti yoki “oziq-ovqat” konsepti ekanini anglash uchun konseptual belgilash foyda beradi.

Xulosa qilib aytganda, ushbu ish tildagi eng ko'p ishlatiladigan otlarning konseptual xaritasini ilk bor eksperimental ravishda tuzdi. Natijalar shuni ko'rsatadiki, o'zbek tili leksik tizimining yadrosini asosan universal konseptlar tashkil etadi, lekin o'ziga xos semantik bo'shliqlar ham mavjud. Tadqiqot jarayonida ko'p ma'nolilik, tarjima va moslashtirishdagi muammolar aniqlandi. Kelgusida konseptual xaritani yanada mukammallashtirish uchun korpuslarni kengaytirish, avtomatik ma'no ajratish vositalarini joriy etish va o'zbek tiliga xos konseptlar bazasini boyitish zarur. Ushbu yo'nalishdagi izlanishlar o'zbek tilining raqamli sohadagi rivoji va ko'p tilli lingvistik hamkorlik uchun mustahkam zamin yaratadi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. *O'zbek tili elektron korpusining kompyuter modellari*. – GlobeEdit nashriyoti, 2021. – 148 b. (Monografiya).
2. Abjalova M. Lingvistik kompyuter dasturlari uchun ma'lumotlar bazasi va lingvistik protsessor. // “O'zbek tilini dunyo miqyosida keng targ'ib qilish bo'yicha hamkorlik istiqbollari” mavzusidagi xalqaro ilmiy-amaliy anjuman materiallari. – 2020 yil 19-20 oktabr. – B. 509-512.
3. Agostini A., Usmanov T., Xamdamiy U., Abdurahmonova N., Mamasaidov M. UzWordNet: A Lexical-Semantic Database for the Uzbek Language. // Proceedings of the 11th Global Wordnet Conference. – 2021. – P. 8-19.
4. List J.-M., Cysouw M., Forkel R. Concepticon: A Resource for the Linking of Concept Lists. // Proceedings of the 10th Int. Conf. on Language Resources and Evaluation (LREC). – Portorož, Slovenia: European Language Resources Association, 2016. – P. 2393-2400.
5. Tjuka A., Forkel R., List J.-M. Curating and extending data for language comparison in Concepticon and NoRaRe. – Open Research Europe, 2023. – 2:141 (<https://doi.org/10.12688/openreseurope.15380.2>).
6. Miller G.A. WordNet: a lexical database for English. – Communications of the ACM, 1995. – Vol. 38, No. 11. – P. 39-41.
7. Abjalova M.A. Lingvistik ontologiya – leksik ma'lumotlar bazasi. // *O'zbekiston: til va madaniyat* (Ilmiy jurnali, ToshDO'TAU). – 2021. – №4. – B. 40-60.

O'ZBEK TILI TA'LIMIY KORPUSI YARATILISHI, UNING TUZILISHI, IMKONIYATLARI VA KAMCHILIKLARI

Hakimova Nargiza

Mirzo Ulug'bek nomidagi

O'zbekiston Milliy universiteti

Jurnalistika va o'zbek filologiyasi fakulteti

E-mail: nargizashavkatovna.07@gmail.com

Annotatsiya: Ushbu maqolada o'zbek tilining ta'limiy korpusi tuzilishi, imkoniyatlari ko'rib chiqilib, uni yaratish tamoyillari va metodologiyasi o'rganiladi. Korpusdan foydalanish jarayonida yuzaga kelishi mumkin bo'lgan muammolar batafsil o'rganilib, ularni bartaraf etish yo'llari yoritiladi.

Kalit so'zlar: milliy korpus, korpus lingvistikasi, interfeys, qidiruv tizimi, lug'atlar, normalizatsiya.

Abstract: This article examines the structure and capabilities of the educational corpus of the Uzbek language, studies the principles and methodology of its creation. The problems that may arise in the process of using the corpus are studied in detail, and ways to overcome them are described in detail.

Keywords: national corpus, corpus linguistics, interface, search engine, dictionaries, normalization.

Аннотация: В статье рассматриваются структура и возможности учебного корпуса узбекского языка, изучаются принципы и методика его создания. Подробно изучаются проблемы, которые могут возникнуть при использовании корпуса, и подробно объясняются способы их устранения.

Ключевые слова: национальный корпус, корпусная лингвистика, интерфейс, поисковая система, словари, нормализация.

Zamonaviy til texnologiyalarining rivojlanishi, sun'iy intellekt asosidagi tizimlarning keng qo'llanilishi kompyuter lingvistikasi sohasida yangicha yondashuvlarni talab qilmoqda. Bu borada lingvistlarning tilni chuqur o'rganish va uni avtomatlashtirish uchun qilgan harakatlari natijasi o'laroq milliy korpus yoki ta'limiy korpusning yaratilishini misol sifatida keltirish mumkin. Ma'lumki, til korpuslari tilni diaxron va sinxron aspektda o'rganish, tilning so'z boyligini oshirish va ularning lug'aviy ma'nosi bilan tanishtirishda katta ahamiyatga ega. Ta'limiy korpus esa yuqorida sanab o'tilgan vazifalarni bajarish uchun aynan qo'l keladi. Shu sababli ham o'zbek tilining ta'limiy korpusi "O'zbek tili Milliy korpusini yaratish konsepsiyasi"ning ilk yirik bosqichi hisoblanadi. Ta'limiy korpus integrallashgan veb-dastur tarzda shakllantirilib, foydalanuvchilarga faoliyatlarini samarali amalga oshirish uchun o'quv yurtining taqsimlangan axborot tizimiga murojaat qilish va ulardan foydalanish imkonini beradi. Ta'limiy korpus – maktab darsliklari va o'quv lug'atlariga asoslangan til korpusi bo'lib, o'zbek tili Milliy korpusining xususiy (tarkibiy) turi hisoblanadi. O'zbek tilining ta'limiy korpusi AM-FZ201908172 raqamli "O'zbek tili ta'limiy korpusini yaratish" mavzusidagi amaliy loyiha doirasida yaratildi, 2021-yilning 23-aprelida dastlabki taqdimoti o'tkazildi, hozirda uning bazasi boyitilmoqda va takomillashtirilmoqda. Mazkur korpus Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universitetida L.Raupova rahbarligidagi loyihaning amaliy natijasi hisoblanadi. [Abjalova M. Korpus lingvistikasi //Uslubiy qo'llanma. Toshkent: Bookmany print. – 2022.]

1) Lug'atlar – o'zbek tilidagi so'zlarning izohlari, ularning omonim, sinonim, paronim va antonimlari, bundan tashqari ularning uslubiy xoslanishi, bo'yoqdorligi, darajalanish qatori bor yoki yo'qligi haqida ham to'liq ma'lumot olish mumkin. Bu so'zlar ichida ko'p qo'llanidiganlari

alohida ajratib qo'yilgan. (Faol so'zlar – 8786ta) Kiritilgan so'zlardan 3284tasining audio shakli ham bor.

2) Qidiruv tizimi – so'z yoki so'z brikmasini qidirib, undan qaysi asarlarda yoki matnda foydalanilganini bilish va aynan shu so'z yoki so'z brikmasi qatnashgan gaplarni ko'rish imkoniyatini beradi.

3)Xorijliklar – xorijliklar uchun o'zbek tilini o'rganishda qo'l keladigan qiziqarli videolar joylangan bo'lib, hozirda ular 35tani tashkil qiladi.

Ta'limiy korpus yaratish jarayonida bir necha metodologik bosqichlarga asoslanilgan:

- 1) **Matnni yig'ish va tanlash.** Korpus asosan o'quv darsliklari, ilmiy maqolalar va boshqa o'quv materiallaridan iborat;
- 2) **Tozalash va normalizatsiya.** Matnni tozalash va normalizatsiyalash jarayonida noto'g'ri yozilgan so'zlar, orfografik xatolar to'g'rilangan;
- 3) **Belgilash (annotatsiyalash).** Morfologik, sintaktik yoki semantik tahlil orqali matnga teglar qo'yilgan;
- 4) **Korpus arxitekturasini ishlab chiqish.** Ma'lumot bazasi va interfeysi tuzilgan;
- 5) **Sinov va qo'llash:** Korpus til o'rgatishda, analizda va tadqiqotlarda sinovdan o'tkazilgan.

O'zbek tili ta'limiy korpusi o'z foydalanuvchilari uchun bir qator imkoniyatlar yaratishi bilan birga unda yetarlicha kamchiliklar ham bor. Bular ichida eng sezilarlisi resurslar sonining – mavjud matnlar bazasining cheklanganligi hisoblanadi. Korpus asosan maktab darsliklari va o'quv lug'atlaridan tashkil topgan. Bu esa tilning uslubiy va janr jihatidan xilma-xilligini to'laligicha aks ettira olmaydi. Shu sababli qidiruv tizimining cheklanganligi va statistik tahlil imkoniyatlarining yetarli emasligi kuzatiladi. Bundan tashqari korpusda o'zbek tili frazemalari bazasining yetishmasligi ham katta muammolardan biri. Ma'lumki, o'zbek tili o'zining turli iboralari bilan mashhur. Badiiy asarlarni, umuman, o'zbek tilini to'laligicha tushunish uchun, albatta, iboralarning ma'nosini ham bilish talab qilinadi. Chunki deyarli barcha o'zbek yozuvchilari asarlarida iboralar mavjud bo'lib, ular asar mazmunini yanada boyitishga xizmat qilgan. Bundan tashqari so'zlashuv uslubida ham frazeologizmlardan keng foydalaniladi. Shunday ekan, frazemalar uchun alohida bir bo'lim qilish maqsadga muvofiq bo'lardi, albatta. Hozirda ham korpusga iboralar va ularning mazmun-mohiyati haqida ma'lumotlar kiritilgan bo'lib, ammo ularning soni juda kam, ko'pchiligi ma'lumotlar bazasiga hali kiritilmagan. Yuqoridagi kamchiliklarga sabab sifatida o'zbek tili ta'limiy korpusi yaratilganligiga hali ko'p bo'lmaganini keltirish mumkin.

Xulosa qilib aytganda, o'zbek tili o'zining ta'limiy korpusiga ega bo'lganining o'ziyoy uning katta yutuqlaridan biri hisoblanadi. Chunki ta'limiy korpus tilni o'rganish, tahlil etish, uni kelajak avlodga to'lig'icha yetkazish va o'qitish jarayonida zamonaviy va samarali vosita sifatida katta ahamiyat kasb etadi. Bunday korpus o'quvchilar va tadqiqotchilar uchun tilni yetarli darajada mukammal o'rganishi uchun boy til materiali, real misollar va statistik tahlil imkoniyatini yaratadi. Shunday ekan, korpusning foydalanuvchilarga beradigan katta imkoniyatlarini yanada rivojlantirish lingvistlarning oldida turgan eng katta maqsadi bo'lishi kerak. O'zbek tili ta'limiy korpusining ayrim jihatlarida yetarlicha kamchiliklar mavjud bo'lib, ularni bartaraf etish uchun hali ko'plab amaliy ishlar olib borish talab etiladi. Hozirda bu borada turli tadqiqotlar olib borilmoqda. Ko'plab tilshunoslar korpusni yanada boyitish maqsadida turli uslub va janrdagi matnlarni kiritish, frazeologik birliklar va lahjaviy xususiyatlarni qamrab olishni taklif qilishmoqda. Bu takliflar amalga oshirilsa, korpus yanada takomillashib, mukammal ko'rinish olgan bo'lardi.

Foydalanilgan adabiyotlar:

1. Абжалова М.Х., Рашидова У.Б. Til korpuslarida frazemalar bazasini yaratish omillari. – ResearchGate – 2023.

2. Алимова С.Ю. О'zbek tilshunosligida korpus lingvistikasi muammolari. // Filologiya masalalari. – 2021. – №3. – Б. 54–59.
3. Тожибоев Ш., Саттарова Д. О'zbek matn korpuslarining lingvistik tahlil imkoniyatlari. // Til va adabiyot ta'limi. – 2022. – №2. – Б. 40–45.
4. Исломова З. Zamonaviy lingvistikada korpus tilshunosligi va uning ahamiyati. // Ilmiy axborotlar. – 2020. – №4. – Б. 30–33.
5. Бектурдиева Н. Korpus lingvistikasi asoslari. – Toshkent: Fan, 2019. – 112
6. Холйоров, У. О'zbek tilida o'quv korpusini yaratish muammolari // Filologiya masalalari. – 2022. – №2. – Б. 45–50.
7. Abjalova, M. X. Korpus lingvistikasi: o'quv-uslubiy qo'llanma. – Toshkent: O'zMU, 2021. – 48 b.

STRUCTURAL MODELS OF TRANSLATION OF UZBEK AND ENGLISH MATHEMATICAL TERMS

Abdurakhmonova Nilufar

Doctor of science, professor
National University of Uzbekistan

Zokirova Khulkar Izzatillayevna

Senior teacher, Angren university

Annotation: This article analyzes structural models used for translating mathematical terminology between Uzbek and English languages. Special attention is given to translation methods such as borrowing, calque, functional equivalence, and descriptive translation. The paper discusses challenges encountered in translating mathematical terms due to linguistic asymmetry and cultural differences. Examples illustrate effective solutions employed by translators to overcome these challenges, highlighting the practical aspects of translation techniques in mathematics-specific contexts.

Keywords: translation models, mathematical terminology, Uzbek-English translation, structural translation methods, linguistic asymmetry.

Аннотация: В статье рассматриваются структурные модели перевода математической терминологии между узбекским и английским языками. Особое внимание уделяется таким методам перевода, как заимствование, калькирование, функциональная эквивалентность и описательный перевод. Обсуждаются сложности, возникающие при переводе математических терминов из-за лингвистической асимметрии и культурных различий. Приводимые примеры показывают эффективные решения, применяемые переводчиками для преодоления этих трудностей, и подчёркивают практическое значение техник перевода в специализированном математическом контексте.

Ключевые слова: модели перевода, математическая терминология, узбекско-английский перевод, структурные методы перевода, лингвистическая асимметрия.

Translating scientific terminology is a crucial yet challenging task in any language pair. In the context of English–Uzbek and Uzbek–English translation of mathematical terms, translators must accurately convey specialized concepts while accounting for linguistic and cultural differences. Over the 20th century, Uzbek mathematical terminology developed through deliberate translation efforts, drawing on sources in Arabic, Russian, and international English lexicons. “The accurate rendering of technical terms in Uzbek requires syntactic and morphological alignment with the source language, particularly in translating sentence-level units.” [Abduraxmonova N. Z., 2018: 14]

Many mathematical terms in Uzbek today have either been **calqued** (literally translated) from foreign terms or **borrowed** directly. Yet, gaps remain where no precise equivalent exists, requiring creative solutions by translators. This section examines the established methods for translating terminology – with a focus on mathematics – and analyzes how they are applied in English–Uzbek and Uzbek–English contexts. It discusses classical translation models as a framework, then classifies practical translation methods (borrowing, calque, functional equivalents, descriptive translation, etc.) with examples. Challenges such as asymmetry of terminologies (when one language lacks a term present in another) and strategies to resolve them are also explored. “Lexical units, especially mathematical terms, often require morphological parsing before translation to avoid incorrect equivalency in Uzbek.” [Abdurakhmonova N., 2016: 15] The goal is to identify effective strategies for translating mathematical terms and highlight persistent issues that translators and terminologists face.

Terminology translation in mathematics or technical demands precision and consistency. “Technical terms in Uzbek verbs often consist of analytic constructions, which must be modeled for correct semantic transfer in machine translation.”[Abdurakhmonov N., 2017: 93] Unlike general vocabulary, terms carry specific definitions; a mistranslation can mislead or confuse learners and professionals. Some English terms, especially those coined in recent decades, have no established Uzbek equivalent, resulting in either adoption of the English term (often via Russian) or the creation of a new term. Conversely, some Uzbek mathematical expressions may be opaque to an English reader if translated word-for-word (for instance, “**tub son**” literally means “root number” but actually means “**prime number**”).

Modern translation studies provide various models that classify translation techniques. A classic framework is offered by J.Vinay and J.Darbelnet, who distinguish between **direct** (literal) and **oblique** (free or structural) translation methods [Vinay J.-P., Darbelnet J., 1995]. Direct methods include **borrowing**, **calque**, and **literal translation**, which involve minimal alteration of the source text structure. Oblique methods (such as transposition, modulation, equivalence, adaptation) are used when direct translation is impossible due to structural or conceptual gaps. In terminology translation, direct methods are often preferred for scientific terms to preserve precision, but oblique strategies may be needed when a term has no ready equivalent. The issue of **lacunae** (gaps) — if a concept exists in the source language but not in the target language, the translator must fill the void with an appropriate corresponding expression to maintain the “overall impression” of the message [Vinay J.-P., Darbelnet J., 1995]. This insight is directly relevant to English–Uzbek mathematical translation, where many modern concepts had no precedent in Uzbek and had to be introduced via translation.

P.Newmark’s model of translation procedures further refines these methods, especially for cultural and technical terms. He identified procedures such as **transference** (direct transfer of a word), **naturalization** (adapting the word to target pronunciation/spelling), **through-translation** (calque), **functional equivalent**, and **descriptive equivalent**, among others [Newmark P., 1988]. Two particularly relevant procedures for terminology are the *functional equivalent* and the *descriptive equivalent*. A **functional equivalent** means using a term or expression in the target language that has an equivalent *function* or role as the source term, often a culture-neutral word. This is useful in technical translation to find a concept in the target language that fulfills the same role as the source concept (e.g. using an existing Uzbek word for a mathematical idea if it conveys the same function or property). A **descriptive equivalent**, on the other hand, is essentially a definition: the meaning of the term is explained in a phrase or sentence in the target language. This method is used when no concise term exists, so the translator provides a description to convey the concept.

In summary, the theoretical frameworks establish a spectrum from direct transfer to explanatory translation. For mathematical terminology, the most pertinent methods are borrowing (transference), calque, functional equivalence, and descriptive translation. In practice, translators decide among these based on the existence of an Uzbek term, the clarity for the target audience, and the need to preserve the precise meaning. We will now focus on each method in turn, showing how they are applied in English–Uzbek/Uzbek–English mathematical translation, supported by examples from textbooks, dictionaries, and corpus data.

Borrowing is the simplest method: taking the source term directly into the target language with minimal changes. In translation of mathematical terminology, borrowing often involves **transliteration** – writing the term in the Uzbek alphabet (Latin script) – since Uzbek will not use the English spelling but preserves the pronunciation. Borrowing is typically used when the concept is new or there is no existing term in Uzbek, or when an international term is widely recognized. Borrowing is a common way to overcome a *lacuna* for an “unknown concept” and is often the first recourse for translators. Many mathematical terms in Uzbek have been adopted this way, especially those of Greek/Latin origin that came via Russian or directly from English.

Some examples of **borrowed mathematical terms** (English → Uzbek) include:

- *matrix* – **matritsa** (матрица)
- *vector* – **vektor** (вектор)
- *algorithm* – **algoritm** (алгоритм)
- *diameter* – **diametr** (диаметр)
- *perpendicular* – **perpendikulyar** (перпендикуляр)

In each case, the English term (often itself of Latin origin) is transliterated into Uzbek. The term “**algoritm**” is especially interesting: it originates from the name of the medieval scholar al-Khwarizmi (whose name in Uzbek is *al-Xorazmiy*), yet modern Uzbek does not use a native coined term but the international form *algorithm*. This reflects how global scientific terms are often uniformly adopted. Similarly, “**vector**” and “**matrix**” were introduced into Uzbek through Russian during the Soviet period; they are now standard Uzbek terms written in the Latin script. The method of borrowing via transliteration is characterized by preserving the original term’s form as much as possible, letter by letter, in the target script. The rationale is to avoid losing the term’s identity or “national character” when no native equivalent exists. [Musaev O., 2005]

Borrowing is prevalent for higher mathematics or specialized branches. For example, terms like “**endomorphism**” or “**homeomorphism**” have no Uzbek analogs and are rendered as *endomorfizm*, *gomomorfizm*, etc. . In a bilingual math dictionary analysis, transliteration alone accounted for roughly 17% of the term renderings. Uzbek textbooks at times include borrowed terms in parentheses after an explanation to familiarize students with international terminology.

A **calque**, or loan translation, is a term or phrase translated component-by-component into the target language. This method creates a new Uzbek term by translating the meaningful parts of the English term using Uzbek words. In mathematical terminology, calquing has been a productive strategy to build the Uzbek lexicon, especially during early language planning efforts. By using existing roots and words, calques produce terms that are transparent to Uzbek speakers. Vinay and Darbelnet describe calque as a special kind of borrowing where the *form* of the source expression is borrowed but translated literally into the target language. For English–Uzbek, descriptive translation might be applied for very new terms or niche concepts not in common use. For example, consider the term “*skew lines*” (lines in 3D space that do not intersect and are not parallel). Uzbek mathematics might not have a fixed term for this (Russian uses “*skreshivayushchiesya pryamye*”). A translator could render this descriptively as “**kesishmaydigan va parallel bo‘lmagan to‘g‘ri chiziqlar**” – literally “straight lines that do not intersect and are not parallel.” This phrase fully describes the concept of skew lines. If this concept appears only once in a text, a descriptive translation is acceptable and ensures the reader understands it. Another example: “*self-similar set*” (in fractal geometry) might be translated as “**o‘zi o‘ziga o‘xshash to‘plam**”, essentially unpacking the meaning of self-similarity in words since a concise term didn’t exist in standard Uzbek math vocabulary.

Descriptive translations often appear in glossaries or when introducing a term for the first time. A common practice in academic writing is to give a descriptive explanation in the target language and perhaps also mention the original term in parentheses. For instance: “*X is an example of a **butun ko‘rinuvchi sonlar to‘plami** (literally, a set of numbers that appear whole; known in English as an integer lattice).*” Here the translator might not be sure if *integer lattice* has a known Uzbek term, so they describe it and optionally provide the English name for reference. In a pedagogical setting, an instructor might say: “**Nosimmetrik matritsa** – ya’ni, o‘zining transpozitsiyasiga teng bo‘lmagan matritsa.” This means “a non-symmetric matrix – that is, a matrix that is not equal to its transpose.” The part after the dash is a descriptive definition clarifying the term. In written translation, that definition could stand in for the term itself if needed.

Translators sometimes choose a descriptive phrase to avoid ambiguity. For example, early on, instead of adopting “*limit*” as *limit* or coining something, some texts explained it as “**cheksiz kichik qiymatga intiluvchi miqdor**” (a quantity that approaches an infinitely small value) for the concept of limit. Eventually *limit* (borrowed) became the standard term, but the descriptive approach was used in initial expository contexts. It sacrifices brevity for clarity. The translator

must ensure that the description is accurate and unambiguous. One challenge is that a descriptive equivalent can sometimes be *too* generic. For example, translating “**random variable**” descriptively as “**tasodifiy o‘zgaruvchan kattalik**” (a quantity that changes randomly) indeed conveys the meaning, but it’s quite wordy. In practice, a term “**tasodifiy miqdor**” has been used in Uzbek (a semi-calque of Russian *sluchaynaya velichina*). If that term weren’t known, a translator might write a full description.

While descriptive translation ensures nothing is lost in terms of content, it has downsides: the flow of the text can be disrupted by long explanatory insertions, and it does not provide a neat terminological unit that can be reused. If a concept recurs often, it’s inefficient to describe it every time; better to introduce a term (via borrowing or calque) and then use it consistently. Therefore, translators and terminologists prefer to create a term (by borrowing or calque) if the concept will be frequently referenced. Descriptive translation is more of a stopgap or explanatory tool.

In Uzbek–English translation, descriptive translation might be employed if an Uzbek term is very local or obscure. For example, “**tub son**” might initially be rendered as “*a prime number (literally ‘root number’ in Uzbek)*” so that the English reader isn’t confused by the literal translation. Or “**ko‘pburchak**” (polygon, literally “multi-angle”) could be glossed as “*polygon (figure with many sides)*” in a translated text if the context requires clarification. Thus, paraphrase can go in either direction to bridge understanding.

We listed some examples showed inappropriate translation in Google translate.

English	Google translate	Appropriate translation
Reciprocal identities	O‘zaro identifikatsiyalar	Teskari ayniyatlar
Algebraic identities	Algebraik identifikatsiyalar	Algebraik ayniyatlar
Relatively prime	Nisbatan eng yaxshi	O‘zaro tub sonlar
Skip count	Hisobni o‘tkazib yuborish	Bitta tashlab sanash(1.3.5)
Tally table	Uzun stol	Hisob jadvali
Terminating decimals	Tugatuvchi o‘nli kasrlar	Chekli o‘nli kasrlar
Thousands	Minglab	Mingdan bir
Times table	Vaqt jadvali	Karra jadvali
Twin primes	Egizak primalar	Ketma-ket tub sonlar
term of sequence	ketma-ketlik muddati	Ketma ket had
Unlike denominators	Denominatorlardan farqli o‘laroq	Har xil maxrajli kasrlar
Unlike numerators	Numeratorlardan farqli o‘laroq	Har xil suratli kasrlar
Zero product property	Nol mahsulot mulki	0 ko‘paytirish xossasi
Base ten number	Asos o‘n raqam	O‘nlik raqamlar

In practice, translators often use a combination of the above methods to render a term. For example, a translator might both transliterate a term *and* give a descriptive explanation: “**graph (grafik) funksiyasi**” or “**grafik (chizma) tasvir**”, combining the borrowed word with a clarifying Uzbek word. This ensures the reader learns the foreign term but also understands it. In the English–Uzbek mathematics dictionary analyzed by Ruzmetova, a small percentage of terms (around 5%) were presented in **both calqued and transliterated forms together**– essentially providing synonyms. For instance, the entry for “*binomial*” listed “**ikki hadli**” and “**binominal**” as Uzbek equivalents. Here *ikki hadli* is a full calque (meaning “two-term”) and *binominal* is a direct borrowing of the international term. By offering both, the dictionary acknowledges usage of both an Uzbek term and an international term in practice. Similarly, “*configuration*” had “**shakl**” (shape) and “**konfiguratsiya**”, “*equivalent*” had “**teng**” and “**ekvivalent**”, etc.. This dual listing reflects an ongoing terminological ambivalence or transition: sometimes the borrowed term is widely used among specialists, while a native equivalent exists or is encouraged for didactic purposes. Translators might choose one or the other depending on context, or even use the combination (e.g., introducing a term as “*teng (ekvivalent)*” to cover both bases).

Overall, combined and auxiliary strategies ensure that the translation of terminology is not only accurate but also usable in context. A translator might start with one method (say, calque) and then adjust by adding a borrowed term in parentheses, or decide to borrow but add a short description, etc. The choice is guided by the **target audience's knowledge** and the **purpose** of the text. In a textbook, clarity is paramount – thus more explanation and native terms. In a research article, conciseness and adherence to international terms might take priority – thus more borrowing and accepted equivalents, with fewer explanatory additions.

Before moving to the conclusion, it's worth noting that Uzbek linguists have actively evaluated these methods in recent years. Methods of translating culturally-loaded words and notes the importance of maintaining the “national character” of terms where possible. Although mathematics is culturally neutral, the sentiment extends to preserving the Uzbek language's integrity by not needlessly importing foreign words when an Uzbek formulation can be used [Hakimov H. I., 2019: 38-42]. Meanwhile, *Fundamentals of Translation Theory* emphasizes choosing the appropriate method to avoid “depriving the translation” of the nuances present in the original. [Musaev O., 2005]

A clear example of asymmetry is the term “*calculus*.” English distinguishes *Calculus* (as in the field of calculus, encompassing differentiation and integration). Uzbek does not have a single native word for this branch of mathematics; often it's referred to by descriptive means (such as “**differensial va integral hisob**”, literally “differential and integral calculation”) or just borrowed as “**analiz**” (from Russian “matematicheskii analiz”). Here, English has one word while Uzbek needs a phrase or borrowed term. The translator must decide whether to borrow (*analiz*) or describe. In academic texts, “**matematik analiz**” is commonly used, a half-borrowing (analysis) with an Uzbek adjective, illustrating how asymmetry is resolved by partial borrowing.

Effective strategies identified include favoring calque or functional translation for fundamental terms to maintain Uzbek readability (especially in educational materials), and using borrowing for highly specialized or internationally standardized terms to maintain consistency with global science. In educational settings, pairing a borrowed term with a descriptive or calqued explanation has proven useful in teaching new concepts without overwhelming students with foreign vocabulary. Meanwhile, terminologists have worked to standardize many terms, reducing confusion – the presence of dictionaries and corpus-based research helps translators choose the right term among alternatives.

Despite this progress, **persistent challenges** remain. Terminological asymmetry means translators occasionally must invent or agree upon new Uzbek terms for emerging concepts (for example, new developments in computer science or advanced mathematics). Ensuring these new terms are disseminated and accepted is an ongoing effort involving translators, subject experts, and language authorities. Inconsistent usage is still an issue: different sources might use different Uzbek terms for the same English concept, which can hamper mutual understanding. Additionally, the rapid growth of English as the lingua franca of science puts pressure on languages like Uzbek to continuously update their terminologies; translators often serve as the first line in that updating process. They need to keep abreast of both languages' evolving usage, sometimes even proposing neologisms or advocating for one translation method over another in academic forums.

In conclusion, the translation of mathematical terminology between English and Uzbek is a dynamic interplay of linguistic creativity and adherence to scientific precision. By applying a classification-based approach to translation methods, translators can systematically tackle each term: deciding whether to transfer it unchanged, translate its parts, find an analogous term, or explain it outright. The examples and analysis provided show that while many English mathematical terms now have well-established Uzbek renditions, the translator's role in refining and expanding this terminological repertoire remains vital. As Uzbek continues to develop as a language of science and education, the effective strategies outlined – judicious borrowing, meaningful calques, clear functional equivalents, and, when needed, concise descriptions – will guide translators in bridging languages without losing meaning. The challenges of asymmetry and

standardization are being addressed step by step, and each successful translation of a term is a small victory in making mathematical knowledge accessible to learners and professionals across languages.

References:

1. Abdurakhmonov N. Modeling Analytic Forms of Verb in Uzbek as Stage of Morphological Analysis in Machine Translation // Journal of Social Sciences and Humanities Research. – 2017. – Vol. 5, № 3. – P. 89–100. –
2. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation // Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. – 2016. – № 2 (38). – P. 12–17
3. Abdurakhmonova N., Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation // Foreign Philology: Language. Literature, Education. – 2018. – № 3. – P. 68–75.
4. Abdurakhmonova N., Urdishev K. Corpus based teaching Uzbek as a foreign language // Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL). – 2019. – Vol. 6, № 1. – P. 131–137.
5. Abduraxmonova N. Z. Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) dissertation abstract. – Tashkent, 2018. – 42 p.
6. Akmalov A. The founder of Uzbek mathematical terminology // Ma'rifat. – 2007. – Sept. 12. – P. 7.
7. Aripov M., Sharipbay A., Abdurakhmonova N., Razakhova B. Ontology of grammar rules as example of noun of Uzbek and Kazakh languages // Abstracts of the VI International Conference “Modern Problems of Applied Mathematics and Information Technology – Al-Khorezmiy 2018”. – Tashkent, 2018. – P. 37–38.
8. Hakimov H. I. Methods of translating words denoting national values // Foreign Philology. – 2019. – № 2. – P. 38–42.
9. Musaev (Musayev) O. Fundamentals of Translation Theory. – Tashkent: Fan, 2005. – 208 p.
10. Newmark P. A Textbook of Translation. – New York: Prentice Hall, 1988. – P. 292
11. Ruzmetova O. A. Study of Mathematical Terms in Bilingual Dictionaries // Journal of Applied Linguistic Studies. – 2022. – Vol. 9, № 1. – P. 90–95.
12. Sharopova R. J. Semantic-structural features and lexicographic interpretation of socio-political terms in the Uzbek language: PhD dissertation. – Karshi: Karshi State University, 2019. – P.176
13. Sobirova G., Artikova Y., Pidaeva Sh., Djumabaeva I. English–Russian–Uzbek Dictionary (for Mathematics Majors). – Tashkent: Uzbekistan National University, 2004. – P. 325
14. Vinay J.-P., Darbelnet J. Comparative Stylistics of French and English: A Methodology for Translation / transl. J. Sager, M.-J. Hamel. – Amsterdam: John Benjamins, 1995. –P. 358.
15. Kubedinova L., Khusainov A., Suleymanov D., Gilmullin R., Abdurakhmonova N. First Results of the TurkLang-7 Project: Creating Russian-Turkic Parallel Corpora and MT Systems // In: Proceedings of the Computational Models in Language and Speech Workshop (CMLS 2020) co-located with the 16th International Conference on Computational and Cognitive Linguistics (TEL 2020). – 2020. – P. 90–101.
16. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus // In: Proceedings of the International Conference on Language Technologies for All (LT4All). – Paris: UNESCO Headquarters, 2019. – P. 1–6.

O‘ZBEK TILI MATNLARINI STEMLASH ALGORITMLARI

Sharipov Maqsud Siddiqovich

Texnika fanlari nomzodi, dotsent

Abu Rayhon Beruniy nomidagi Urganch Davlat Universiteti

E-mail: maqsbek72@gmail.com

Sattarova Surayyo Beknazarovna

Urganch RANCH Texnologiya Universiteti magistranti

E-mail: surayyosattarova20@gmail.com

Annotatsiya: Ushbu ishdada tabiiy tilni qayta ishlash sohasidagi muhim tushunchalardan biri — stemming, ya’ni so‘zlarni o‘zining o‘zak shakliga keltirish jarayoni haqida fikr yuritiladi. Matnni avtomatik tahlil qilish, qidiruv tizimlari, tarjima vositalari, hujjatlarni tasniflash kabi ko‘plab sohalarda stemming muhim vosita bo‘lib xizmat qiladi. Ingliz, rus, nemis kabi tillar uchun turli algoritmlar mavjud bo‘lsa-da, o‘zbek tili uchun hali yetarlicha samarali stemmerlar ishlab chiqilmagan. Shu sababli, ushbu tezisda Snowball algoritmgiga asoslangan o‘zbekcha stemmer yaratish g‘oyasi nazariy jihatdan yoritiladi. Stemming algoritmining tokenizatsiya va lemmatizatsiyadan farqli jihatlari ko‘rsatib o‘tiladi, uning o‘zbek tilidagi qo‘llanilish sohalari — masalan, chat botlar, avtomatik tarjima, matn tahlili, ijtimoiy tarmoqlarda kuzatuv va raqamli lug‘atlar yaratish kabi yo‘nalishlarda tutgan o‘rni izohlanadi. Shuningdek, kelgusida o‘zbek tili uchun samarali stemmerlar ishlab chiqish va ularni sun‘iy intellekt tizimlariga integratsiya qilish istiqbollari muhokama qilinadi.

Kalit so‘zlar: stemming, o‘zbek tili, Snowball algoritmi, tabiiy tilni qayta ishlash, so‘zlarni normallashtirish, matn tahlili.

Аннотация: В данной тезисной работе рассматривается одна из ключевых концепций обработки естественного языка — stemming, то есть процесс приведения слов к их корневой форме. Stemming служит важным инструментом в таких областях, как автоматический анализ текста, поисковые системы, переводческие программы, классификация документов. Несмотря на наличие эффективных алгоритмов stemming для английского, русского и немецкого языков, для узбекского языка до сих пор не создано достаточно действенных решений. Поэтому в данной работе теоретически обосновывается идея создания узбекского stemмера на основе алгоритма Snowball. Показаны отличия алгоритма stemming от токенизации и лемматизации, а также рассмотрены области его применения в узбекском языке — такие как чат-боты, автоматический перевод, анализ текстов, мониторинг социальных сетей и создание цифровых словарей. Также обсуждаются перспективы дальнейшей разработки узбекских stemмеров и их интеграции в системы искусственного интеллекта.

Ключевые слова: stemming, узбекский язык, алгоритм Snowball, обработка естественного языка, нормализация слов, анализ текста.

Abstract: This thesis discusses one of the key concepts in natural language processing — stemming, which is the process of reducing words to their root form. Stemming plays a crucial role in areas such as automatic text analysis, search engines, translation systems, and document classification. While there are effective stemming algorithms for English, Russian, and German, there is still a lack of efficient stemmers for the Uzbek language. Therefore, this paper theoretically explores the idea of creating an Uzbek stemmer based on the Snowball algorithm. The differences between stemming, tokenization, and lemmatization are highlighted, and various application areas in Uzbek — such as chatbots, machine translation, text analysis, social media monitoring, and digital dictionary creation — are discussed. The thesis also addresses the future prospects of

developing high-performance stemmers for Uzbek and integrating them into artificial intelligence systems.

Keywords: stemming, Uzbek language, Snowball algorithm, natural language processing, word normalization, text analysis.

Kirish. Bugungi raqamli asrda tabiiy tilni qayta ishlash (Natural Language Processing — NLP) texnologiyalari kundalik hayotimizning ajralmas qismiga aylanib bormoqda. Internetdagi qidiruv tizimlari, avtomatik tarjima xizmatlari, virtual yordamchilar, elektron hujjatlarni saralash tizimlari, ijtimoiy tarmoqlardagi matnlarni avtomatik tahlil qilish kabi ko‘plab sohalarda NLP keng qo‘llanilmoqda. Ushbu soha doirasida muhim vazifalardan biri bo‘lgan so‘zlarni o‘zak shaklga keltirish, ya’ni stemming texnologiyasi, matnni tahlil qilishda va uni kompyuterga tushunarli shaklga keltirishda muhim o‘rin tutadi.

Stemming – bu matndagi so‘zlarni ularning asosiy yoki o‘zak shakliga keltirish jarayonidir. Bu jarayon tabiiy tilni qayta ishlash (Natural Language Processing – NLP) tizimlarida so‘zlar o‘rtasidagi semantik (ma’nodosh) aloqalarni aniqlash va tushunish imkonini beradi. Oddiy qilib aytganda, stemming – so‘zning turli grammatik shakllaridan (masalan, ko‘plik, egalik, kelishik qo‘shimchalari) yasalgan shakllarini birlashtirib, ularni yagona o‘zak shaklga keltirishdir.

Masalan:

- “boradi”, “bordi”, “boryapti”, “borayotgan” so‘zlarining barchasi “bor” o‘zak shakliga ega.
- “ishchi”, “ishlamoqda”, “ishga”, “ishlar” – bularning o‘zagi “ish” bo‘lib, stemming orqali ularni bitta ma’noda qayta ishlash mumkin bo‘ladi.

Demak, *stemming* — bu tabiiy til so‘zlarini grammatik qo‘shimchalardan tozalab, ularning boshlang‘ich (o‘zak) shakliga keltirish jarayoni hisoblanadi. Bu jarayonning asosiy maqsadi — bir xil ma’noni ifodalovchi, lekin turli grammatik shakllarda ishlatilgan so‘zlarni yagona shaklga keltirishdir. Masalan, ingliz tilida “connect”, “connected”, “connection”, “connecting” kabi so‘zlar barchasi bir o‘zak — “connect” atrofida hosil bo‘lgan. Stemming yordamida ushbu shakllar yagona “connect” o‘zagiga keltiriladi. Bu jarayon qidiruv tizimida foydalanuvchining “connect” deb bergan so‘rovi bo‘yicha “connected” yoki “connection” so‘zlari qatnashgan hujjatlarni ham aniqlash imkonini beradi.

O‘zbek tilida ham so‘zlar ko‘plab qo‘shimchalar orqali yasaladi. Masalan, “yozdi”, “yozyapti”, “yozmoqchi”, “yozgan”, “yozuvchi”, “yozilgan”, “yozmaslik” kabi so‘zlar barchasi “yoz” o‘zagidan hosil bo‘lgan. Shunday bo‘lsa-da, hozirgi kungacha o‘zbek tiliga moslashtirilgan va keng miqyosda amalda qo‘llanilayotgan stemming algoritmlari juda kam. O‘zbek tili agglutinativ til bo‘lgani bois, unda so‘z yasalishi murakkab morfologik tizimga ega bo‘lib, bu stemmer yaratish jarayonini yanada murakkablashtiradi. Agglutinatsiya xususiyati tufayli bitta so‘zga bir nechta grammatik va leksik qo‘shimchalar ketma-ket qo‘shilishi mumkin: “borayotgandirsizmi”, “o‘qitilmaganligimizdan” kabi shakllar bunga misoldir.

Hozirgi kunda ingliz, nemis, fransuz, rus, turk va boshqa ko‘plab tillar uchun yaratilgan Porter, Lancaster, Snowball, Lovins kabi stemming algoritmlari mavjud bo‘lsa-da, ularning barchasi o‘zbek tilining morfologik tuzilishiga to‘liq mos kelavermaydi. Shu bois, ushbu ilmiy ishda aynan Snowball algoritmi asosida o‘zbek tiliga xos qoidalarga moslab stemmer ishlab chiqish taklif etiladi. Snowball algoritmi — bu Porter algoritmining soddalashtirilgan, modul tuzilishga ega va turli tillar uchun qoidalarga asoslangan holatda kengaytirilgan varianti bo‘lib, u turli tilda ishlov beruvchi stemmerlarni yaratishga imkon beradi.

Ushbu tadqiqot ishi uchun o‘zbekcha matn korpusi ustida ishlanib, o‘zbek tilidagi asosiy so‘z yasovchi va grammatik qo‘shimchalar tahlil qilinishi rejalashtirilgan. Keyin esa ular asosida Snowball vositasi orqali stemmer algoritmi ishlab chiqiladi va u test so‘zlarda sinovdan o‘tkaziladi. Olingan natijalar tahlil qilinib, stemming jarayonining aniqligi baholanadi.

Mazkur tadqiqotning dolzarbligi shundaki, u o'zbek tilida avtomatik morfologik tahlil, hujjatlarni indekslash, qidiruv tizimlari, matn tozalash va ma'lumotlarni avtomatik saralash kabi ko'plab sohalarida zamonaviy texnologiyalarni rivojlantirishga xizmat qiladi.

Stemming va uning ahamiyati. Stemming — bu axborotni izlash jarayonini yaxshilashi mumkin bo'lgan jarayonlardan biridir. Stemming degani — so'zning infleksion yoki derivatsion (yasalgan) shakllarini uning o'zak (asosiy) shakliga keltirishdir. Stemming matnni tahlil qilish (text mining) va tabiiy tilni qayta ishlash (natural language processing) jarayonlarida ham qo'llaniladi. Har bir tilning o'ziga xos xususiyatlari va grammatik qoidalari mavjud bo'lganligi sababli, bitta stemming algoritmini barcha tillarga qo'llash juda qiyin.[1].

O'zbek tili agglutinatив til bo'ganligi sababli, ko'plab so'zlar qo'shimchalar qo'shilishi orqali hosil qilinadi va qo'shimchalar soni ham juda ko'p. Shu sababli, so'zlarning o'zak (asosiy) shaklini aniqlash qiyin bo'ladi [2]. O'zbek tilida bir xil shaklda yoziladigan, ammo ma'nosi har xil bo'lgan qo'shimchalar ham mavjud. Shuningdek, ikki yoki undan ortiq qo'shimchaning yonma-yon qo'shilishi orqali hosil bo'lgan murakkab qo'shimchalar ham uchraydi. Masalan, –chi va –lik qo'shimchalari mavjud, lekin morfemalarni ajratish qoidasiga ko'ra, –chilik kabi monolit shakldagi (yagona) qo'shimcha ham mavjud. –chilik qo'shimchasini tahlil qilganda, uning mazmuniga qarab bu qo'shimcha ikki alohida qo'shimchaning birikmasi ([–chi] + [–lik]) yoki bitta monolit qo'shimcha (–chilik) sifatida qaralishi mumkin. Shuningdek, uchta qo'shimchaning ketma-ket qo'shilishi orqali hosil bo'lgan shakllar ham mavjud: ([–gar] + [–chi] + [–lik], odamgarchilik). Masalan: gul+[–chi]+[–lik] → gulchi [gul sotuvchi], gulchilik [gul yetishtirish sohasi]; dehqonchilik (bu yerda "dehqonchi" degan so'z noto'g'ri hisoblanadi) [3]. Chunki, agglutinatив tillarda so'z shakllarining mumkin bo'lgan kombinatsiyalari soni nazariy jihatdan cheksizdir. Bu esa, o'z navbatida, lug'atda mavjud bo'lmagan so'zlarning so'z turkumini aniqlash (POS tagging – so'z turini belgilash) muammosini yuzaga keltiradi [4].

Stemming, tokenizatsiya va lemmatizatsiya farqlari: NLP tizimida matnni tahlil qilish bir necha bosqichdan iborat bo'lib, ular bir-biriga yaqin bo'lsa-da, turli maqsadlarga xizmat qiladi:

Tokenizatsiya – matnni so'z, gap yoki belgilar kabi kichik birliklarga ajratish jarayoni. Masalan, “Bugun havo iliq.” jumlasini tokenizatsiya qilinganda quyidagi birliklarga bo'linadi: [“Bugun”, “havo”, “iliq”].

Stemming – har bir tokenni (so'zni) grammatik qo'shimchalardan tozalab, uni o'zak shakliga keltiradi [7,8,9]. Masalan, “yashayotgan” → “yasha”, “kitoblari” → “kitob”.

Lemmatizatsiya – bu ham o'zak shaklini aniqlashga xizmat qiladi, biroq grammatik va lug'aviy tahlil asosida har bir so'zning to'liq morfologik shaklini aniqlaydi [6]. Lemmatizatsiya natijasida “keldim”, “kelayotgan”, “kelgan” kabi so'zlar hammasi “kelmoq” lemmaga keltiriladi. Stemming esa ularni “kel” shakliga qisqartiradi.

Stemming texnologiyasi bir nechta muhim imkoniyatlarni taqdim etadi:

Leksik hajmni kamaytiradi: So'zlarning o'zagi aniqlangach, turli grammatik shakllar yagona birlik sifatida qaraladi, bu esa matnlar ustida ishlov berishda resurs sarfini kamaytiradi.

Ma'nodosh so'zlarni birlashtiradi: Shakli turlicha bo'lsa-da, bir ma'nodagi so'zlar umumlashtiriladi va bu, ayniqsa, statistik tahlil va izlash tizimlarida katta ahamiyat kasb etadi.

Morfologik murakkablikni soddalashtiradi. O'zbek tilidagi agglutinatив (qo'shimchali) tuzilmani oddiylashtirish orqali matnni tahlil qilish jarayonini tezlashtiradi. Lemmatizatsiya va stemming — axborotni izlash, tarjima dasturlari va til korpuslarida eng muhim texnologiyalardan hisoblanadi [10, 11]

Semantik aloqalarni yaxshilaydi: So'zlar o'rtasidagi bog'lanishni aniqlash osonlashadi, bu esa tavsifiya tizimlari, semantik qidiruv va boshqa intellektual tizimlar uchun muhim.

Stemming bugungi kunda bir qator sohalarida keng qo'llanilmoqda:

Qidiruv tizimlarida stemmingning qo'llanilishi: Stemming algoritmlari qidiruv tizimlarida matnlarni samarali izlash va saralashda muhim rol o'ynaydi. Ushbu tizimlar foydalanuvchi tomonidan kiritilgan so'zning faqat tashqi ko'rinishiga emas, balki uning o'zak shakliga e'tibor

qaratadi. Bu yondashuv qidiruv natijalarining aniqligi va keng qamrovini oshiradi. Masalan, foydalanuvchi qidiruv oynasiga “kitobxonlar” degan soʻzni kiritgan boʻlsa, stemming algoritmi bu soʻzning oʻzagini “kitob” deb aniqlaydi va “kitob”, “kitobxon”, “kitobxonlik”, “kitobxonga”, “kitoblarga” kabi turli shakllarini ham qamrab oladi. Shunday qilib, foydalanuvchiga bu soʻzga aloqador kengroq maʼlumotlar bazasi taqdim etiladi.

Agar stemming algoritmidan foydalanilmasa, tizim “kitobxonlar” soʻzini faqat aynan shu shaklda matnda mavjud boʻlsa, izlab topa oladi. Biroq, oʻzak asosida qidirish esa koʻproq mos natijalarni taqdim etadi, chunki u barcha morfologik shakllarni birlashtira oladi. Bu esa, ayniqsa, oʻzbek tilidek aglutinativ (yaʼni qoʻshimchalar orqali soʻz shakllari kengayadigan) tillar uchun nihoyatda muhimdir.

Stemming nafaqat qidiruvni aniq va tezkor qiladi, balki katta hajmdagi matnlarni tahlil qilishda tizim yukini kamaytiradi. Chunki soʻzlar bitta oʻzak ostida guruhlanadi, indekslash ishlari yengillashadi va qidiruv tizimi foydalanuvchining ehtiyojiga eng yaqin natijalarni taqdim eta oladi. Shu sababli, stemming algoritmilarining qidiruv tizimlaridagi oʻrni zamonaviy axborot izlash texnologiyalarida ajralmas element hisoblanadi.

Hujjatlar tasnifida stemmingning qoʻllanilishi: Hujjatlarni tasniflashda, yaʼni ularni maʼlum mavzular, sohalar yoki yoʻnalishlarga ajratishda stemming muhim vosita boʻlib xizmat qiladi. Har bir matnda biror mavzuga oid koʻplab morfologik shakllarda uchraydigan soʻzlar boʻlishi mumkin. Masalan, “sotuvchi”, “sotuvchilik”, “sotuvchilarning”, “sotuvchidan” kabi turli shakllar aslida barchasi bir soʻz – “sotuv” yoki “sot” atrofida shakllangan boʻlib, ular savdo faoliyatini ifodalaydi.

Stemming algoritmi [12] soʻz shakllarini yagona oʻzak – “sot” shakliga keltirganida, ushbu hujjat savdo sohasiga oid ekani aniqlanadi. Shunda hujjatlar avtomatik tarzda toifalarga – masalan, “iqtisodiyot”, “savdo”, “marketing” kabi sohalarga ajratiladi. Bu esa katta hajmdagi matnlar bazasida tez va samarali tasniflash imkonini beradi.

Shuningdek, hujjatlar ichidagi soʻz shakllarining oʻzaklari aniqlanib, ularning statistik tahlili amalga oshiriladi. Qaysi oʻzaklar koʻproq ishlatilgani asosida hujjatning asosiy mavzusi yoki sohasi aniqlanishi mumkin. Natijada, matnlarni saralash, arxivlash yoki foydalanuvchiga mos hujjatni topish jarayonlari avtomatlashtiriladi va tezlashadi.

Avtomatik tarjimada stemmingning ahamiyati: Avtomatik tarjima tizimlarida matnni toʻgʻri va mazmunli tarjima qilish uchun, soʻzlarning grammatik shakllarini aniqlash muhim hisoblanadi. Oʻzbek tilida bir soʻz koʻplab qoʻshimchalar yordamida turli maʼnolar kasb etadi. Masalan, “boradi”, “borgani”, “borayotgan”, “borayapti” kabi shakllarning barchasi bir oʻzak – “bor” soʻziga bogʻliq. Stemming algoritmi ushbu shakllarni “bor” oʻzak soʻziga keltirib, tarjima tizimiga ushbu soʻzning asosiy maʼnosini aniqlashga yordam beradi. Natijada, avtomatik tarjima tizimi soʻzlarning grammatik formalariga aldanmasdan, aniq va mantiqiy tarjima bera oladi. Bu, ayniqsa, aglutinativ (qoʻshimchali) tillarda tarjimaning aniqligini sezilarli darajada oshiradi.

Chatbotlar va ovoqli yordamchilarda stemmingning qoʻllanilishi: Chatbotlar va ovoqli yordamchilar foydalanuvchi bilan muloqot qilganda, aytilgan yoki yozilgan soʻzlarning mazmunini tushunishlari kerak. Oʻzbek tilida bitta maʼno turli grammatik shakllarda ifodalansa-da, ular barchasi yagona oʻzakka ega boʻladi. Misol uchun, “yordam bering”, “yordamchi”, “yordam soʻrayapman” kabi jumlar asosan “yordam” soʻziga bogʻliq. Stemming algoritmi bu soʻz shakllarini tahlil qilib, ularning barchasini “yordam” oʻzak soʻziga keltiradi. Shu orqali chatbot yoki ovoqli yordamchi foydalanuvchining ehtiyojini aniq tushunadi va unga mos javob qaytaradi. Bu texnologiyalarda foydalanuvchi soʻzlarining oʻzak shaklini aniqlash muloqot sifatini oshirishda va notoʻgʻri tushunishlar oldini olishda muhim rol oʻynaydi.

Ijtimoiy tarmoqlarda matn tahlilida stemmingning oʻrni: Bugungi kunda ijtimoiy tarmoqlarda foydalanuvchilar tomonidan yoziladigan post va izohlar katta hajmdagi matnlar oqimini tashkil etadi. Ushbu matnlarni tahlil qilishda kontekstni toʻgʻri aniqlash muhim ahamiyatga ega. Chunki bir fikr turli grammatik shakllarda ifodalangan boʻlishi mumkin. Masalan, foydalanuvchilar “yoqdi”, “yoqmaydi”, “yoqmoqda”, “yoqqan” kabi soʻzlar orqali oʻz fikrini

bildiradi. Stemming algoritmi bu soʻz shakllarini umumiy “yoq” kabi oʻzak shaklga keltirib, ularni yagona maʼno doirasida tahlil qilish imkonini beradi. Bu esa foydalanuvchining hissiy munosabati – ijobiymi yoki salbiymi – aniq aniqlanishiga xizmat qiladi. Shu tariqa, ijtimoiy tarmoq kontentini avtomatik ravishda tahlil qilish, foydalanuvchi fikrlarini umumlashtirish va tendensiyalarni aniqlash imkoniyati oshadi.

Oʻzbek tili uchun stemmingning afzalliklari: Oʻzbek tili aglutinativ til boʻlgani sababli, bir oʻzak soʻzga turli grammatik qoʻshimchalar qoʻshilishi orqali yuzlab yangi soʻz shakllari hosil boʻladi. Masalan, “kel”, “kelmoqda”, “kelgan”, “kelmaslik”, “kelmayapti” kabi koʻplab shakllar aslida bitta oʻzak – “kel” soʻziga asoslanadi. Stemming algoritmi ushbu shakllarni tahlil qilib, barchasini yagona oʻzak shaklga keltiradi. Bu jarayon matndagi turli koʻrinishdagi soʻzlarni birlashtirish orqali lugʻat (vocabulary) hajmini sezilarli darajada kamaytiradi. Natijada, kompyuter xotirasi teжалadi, ishlov berish tezligi ortadi va tahlilning aniqligi oshadi. Ayniqsa, katta hajmdagi matnlar bilan ishlovchi tizimlar uchun bu katta afzallik sanaladi. Oʻzbek tilida stemming algoritmini samarali joriy etish orqali lugʻatlarni optimallashtirish, NLP asosidagi amaliy dasturlarni yanada samarali ishlashiga xizmat qiladi.

Soʻzlar oʻrtasidagi semantik aloqalarni aniqlashdagi yordam: Stemming algoritmining muhim afzalliklaridan yana biri — u turli grammatik shakllarda ishlatilgan soʻzlarni yagona oʻzak shaklga keltirish orqali ular oʻrtasidagi semantik (maʼno jihatidagi) bogʻliqlikni aniqlashni osonlashtiradi. Masalan, “ishlamoq”, “ishladi”, “ishlaydi”, “ishlamayapti” kabi soʻzlar yuzaki koʻrinishda turlicha boʻlsa-da, barchasi bitta maʼno – “ish bajarish” atrofida toʻplangan. Stemming ushbu shakllarni “ishla” oʻzak shakliga keltirib, ular oʻrtasidagi semantik aloqani kompyuterga tushunarli qiladi. Bu esa matnni tematik guruhlariga ajratish, mashinaviy tarjima, intellektual izlash, soʻzlar tarmoqlari yaratish kabi koʻplab NLP (Natural Language Processing) sohalarida qoʻl keladi.

Oʻzbek tilidagi murakkab qoʻshimchalar tizimini soddalashtirish: Oʻzbek tili aglutinativ til boʻlganligi sababli, bitta oʻzak soʻzga juda koʻp grammatik va derivatsion qoʻshimchalar qoʻshilishi mumkin. Bu esa kompyuterda tilni avtomatik tahlil qilish jarayonini murakkablashtiradi. Masalan, “kelmayaptilarmi”, “kelyapmizmi”, “kelmasligimizdan” kabi soʻzlar grammatik jihatdan turlicha tuzilgandir, lekin ularning barchasi bitta oʻzak – “kel” soʻziga tayanadi. Stemming algoritmi bu kabi shakllarni oddiy va universal oʻzakka keltirib, murakkablikni sezilarli darajada kamaytiradi. Bu nafaqat matn tahlilini soddalashtiradi, balki dasturlarning tez va samarali ishlashiga ham xizmat qiladi. Ayniqsa, oʻzbek tilining fonetik oʻzgarishlarga boyligi va variantli qoʻshimchalari hisobga olinsa, stemming texnikasi bu boradagi muhim vositaga aylanadi.

Matnni sintaktik va semantik tahlil qilish imkonini kengaytiradi: Stemming algoritmi matnni chuqurroq va aniqroq tahlil qilishda — ayniqsa, sintaktik va semantik (tuzilish va maʼno) jihatdan — katta imkoniyatlar yaratadi. Har bir soʻzning oʻzak shaklini aniqlash orqali, matndagi jumla tuzilmalari (yaʼni gapning tarkibiy qismlari) va soʻzlar oʻrtasidagi maʼno bogʻliqligi aniqroq aniqlanadi. Bu esa mashinaviy tarjima, soʻz birikmalarini aniqlash, soʻrov javob tizimlari (QA systems) va nutqning mantiqiy tarkibini anglash kabi turli NLP vazifalarini bajarishda muhim ahamiyat kasb etadi. Masalan, "olib kelingan kitob", "kitob olib kelindi", "kitobni olib kelishdi" kabi jumlalarda har xil soʻzlar ishlatilgan boʻlsa-da, stemming ularni yagona semantik ifoda sifatida qayta ishlash imkonini beradi. Natijada, kompyuter matnning asosiy gʻoyasi va maʼlumotini aniq tushunishga yaqinlashadi.

Xulosa

Ushbu tezis ishida Snowball algoritmi asosida oʻzbek tiliga moslashtirilgan stemming tizimini ishlab chiqish zarurati va amaliy ahamiyati tahlil qilindi. Oʻzbek tili agglutinativ til boʻlgani bois, undagi soʻz shakllarining xilma-xilligi matnlar ustida avtomatlashtirilgan ishlov berishni murakkablashtiradi. Shu sababli soʻzlarni oʻz oʻzak shakliga keltiruvchi stemming algoritmlarini ishlab chiqish va takomillashtirish bugungi axborot texnologiyalari davrida dolzarb masalalardan biridir.

Snowball algoritmi soddaligi, ko‘p tillarga moslanuvchanligi va qoidabop yozilishi bilan ajralib turadi. Uni o‘zbek tiliga moslashtirish orqali turli sohalarda – qidiruv tizimlari, matnni avtomatik tahlil qilish, hujjatlarni tasniflash, ijtimoiy tarmoqlarda kayfiyat tahlili, tarjima va ovoqli yordamchi tizimlar kabi ko‘plab NLP ilovalarida foydalanish imkoniyati paydo bo‘ladi.

Mazkur tadqiqotda o‘zbekcha matnlar korpusi asosida tahlil olib borilib, qo‘shimchalar ro‘yxati, tokenlash, stemming va natijalarni baholash bosqichlari amalga oshirildi. Dastlabki natijalar stemming texnologiyasi o‘zbek tilida ham samarali natija berishini ko‘rsatdi, ammo aniqlikni oshirish uchun fonetik va morfologik istisnolarni chuqurroq tahlil qilish, lemmatizatsiya bilan birlashtirilgan yondashuvlarni qo‘llash tavsiya etiladi.

Shunday qilib, o‘zbek tilida samarali stemmer yaratish – bu nafaqat texnik, balki lingvistik yondashuvni ham o‘z ichiga olgan murakkab, ammo istiqbolli yo‘nalish bo‘lib, kelajakda NLP va sun‘iy intellektga asoslangan dasturlar rivojlanishiga xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Jalil, M. M., et al. (2017). The development of the Uzbek stemming algorithm. *Advanced Science Letters*, 23(5), 4171–4174.
2. Sharipov, M., & Yuldashov, O. (2022). Uzbekstemmer: Development of a rule-based stemming algorithm for Uzbek language. *arXiv preprint arXiv:2210.16011*.
3. Sharipov, M., & Salaev, U. (2022). Uzbek affix finite state machine for stemming. *arXiv preprint arXiv:2205.10078*.
4. Boltayevich, E. B., et al. (2023). The problem of POS tagging and stemming for agglutinative languages (Turkish, Uyghur, Uzbek languages). In *2023 8th International Conference on Computer Science and Engineering (UBMK)* (pp. 57–62). IEEE.
5. Abjalova, M., Adali, E., & Adilova, M. (2024). The process of lemmatization and stemming in the automatic morphological analysis of Uzbek texts. In *2024 9th International Conference on Computer Science and Engineering (UBMK)* (pp. 1–6). IEEE.
6. Sharipov, Maksud, and Ogabek Sobirov. "Development of a rule-based lemmatization algorithm through Finite State Machine for Uzbek language." *arXiv preprint arXiv:2210.16006* (2022).
7. Izatovich B. I. Development of a stemming algorithm based on a linguistic approach for words of the uzbek language //E-Conference Globe. – 2021. – C. 195-202.
8. Tukeyev U. et al. Computational Model of Morphology and Stemming of Uzbek Words on Complete Set of Endings //2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE). – IEEE, 2024. – C. 1760-1764.
9. Ismailov A. S., Abdurakhmonova N. The development of Alisher stemmer for Uzbek Language //Science and Education. – 2022. – T. 3. – №. 4. – C. 187-213.
10. Sattarova S. B., Bekchanova F. X., Shermetov A. K. Terminologik lug‘at yaratish texnologiyasi va uning ta‘lim tizimidagi ahamiyati //Academic research in educational sciences. – 2023. – T. 4. – №. 5. – C. 422-434.
11. Madatov K. A., Sattarova S. Creation of a Corpus for Determining the Intellectual Potential of Primary School Students //2024 IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM). – IEEE, 2024. – C. 2420-2423.
12. Sharipov M., Salaev U., Matlatipov G. O‘zbek tili fe‘l so‘z turkumi uchun chekli avtomatlar asosida stemming algoritmini yaratish //Computer linguistics: problems, solutions, prospects. – 2021. – T. 1. – №. 1.

O‘ZBEK TILIDAGI IBORALARNING SINONIMIK DARAJALANISHI

Omarova Durdona

ToshDO‘TAU, kompyuter lingvistikasi mutaxassisligi magistranti

E-mail: omarovadurdona@gmail.com

Annotatsiya: Ushbu maqolada o‘zbek tilidagi iboraviy birliklarning sinonimik xususiyatlari va ularning darajalanishi masalasi tahlil etiladi. Tilshunoslikda iboralarning sinonimik qatlamlarini o‘rganish muhim bo‘lib, ular tildagi ma‘no boyligining, uslubiy rang-baranglikning asosiy ko‘rsatkichlaridan biri hisoblanadi. Maqolada iboraviy sinonimlarning tuzilishi, ularning semantik yaqinligi, stilistik yuklamasi va qo‘llanilish doirasi misollar asosida yoritiladi. Shuningdek, iboralarning sinonimik darajalanishi kontekstda qanday aks etishi va matn mazmuniga qanday ta‘sir ko‘rsatishi haqida tahliliy mulohazalar keltiriladi. Tadqiqot natijalari o‘zbek tilshunosligi va tarjima nazariyasi uchun muhim nazariy va amaliy ahamiyatga ega.

Kalit so‘zlar: ibora, sinonim, sinonimik darajalanish, frazeologik birlik, semantik yaqinlik, kontekst, o‘zbek tili

Abstract: This article explores the synonymic characteristics of phraseological units in the Uzbek language and analyzes their gradation. Studying the synonymic layers of phrases is important in linguistics, as they reflect the richness of meaning and stylistic diversity within a language. The article discusses the structure of synonymous phrases, their semantic proximity, stylistic connotations, and usage scope through concrete examples. It also examines how synonymic gradation is expressed in context and how it affects the overall meaning of the text. The results of this study have significant theoretical and practical value for Uzbek linguistics and translation studies.

Keywords: phrase, synonym, synonymic gradation, phraseological unit, semantic proximity, context, Uzbek language

Аннотация: В данной статье рассматриваются синонимические особенности фразеологических единиц в узбекском языке и проводится анализ их градации. Изучение синонимических слоёв фраз представляет собой важное направление в лингвистике, поскольку они отражают смысловое богатство и стилистическое разнообразие языка. В статье освещаются структура синонимических выражений, степень их семантической близости, стилистическая окраска и сферы употребления на основе конкретных примеров. Также анализируется, как синонимическая градация проявляется в контексте и какое влияние оказывает на смысл текста. Результаты исследования имеют как теоретическое, так и практическое значение для узбекского языкознания и теории перевода.

Ключевые слова: фразеологизм, синоним, синонимическая градация, фразеологическая единица, семантическая близость, контекст, узбекский язык

KIRISH

Tilning ifoda imkoniyatlarini kengaytirishda va fikrni aniq, obrazli yetkazishda frazeologik birliklar, xususan, iboralarning o‘rni beqiyosdir. O‘zbek tili boy va rang-barang frazeologizmlarga ega bo‘lib, ularning ko‘pchiligi o‘zaro ma‘nodosh, ya‘ni sinonimlik munosabatiga ega. Bunday birliklarning sinonimik qatlamlari o‘rganish nafaqat ularning ma‘nodoshligini, balki ularning uslubiy, semantik va kontekstual farqlarini ham tahlil qilish imkonini beradi.

Sinonimik iboralar, ayniqsa, adabiy til uslublarida fikr ifodalanishida mazmuniy noziklikni, emotsional rang-baranglikni yaratishda muhim vosita bo‘lib xizmat qiladi. Ularning darajalanishi esa, bir ma‘noni ifodalovchi iboralarning kuchi, emotsional yuklamasi yoki kontekstdagi vazifasiga qarab farqlanishini anglatadi. Masalan, “qattiq ishlamoq” iborasiga “jonini

jabborga berib ishlamoq”, “qul-elni unutib ishlamoq” kabi sinonimik iboralar mos keladi, ammo ularning stilistik yuklamasi va kuchi har xil.

Tilshunos olimlar tomonidan o‘zbek tilida sinonimik birliklarning o‘rganilishi bo‘yicha bir qator tadqiqotlar amalga oshirilgan. Jumladan, A. Madvaliyev o‘zining izlanishlarida iboraviy sinonimlarning semantik yaqinligi va uslubiy farqlarini chuqur tahlil qiladi [Madvaliyev A,2015].G. Qobulova esa iboraviy birliklarning semantik turlari va ularning kontekstda tutgan o‘rni batafsil yoritib bergan [Qobulova G, 2024].

Shuningdek, “O‘zbek tilining frazeologik lug‘ati” [Mdrahimov N, 2006] va “O‘zbek tili sinonimlarining izohli lug‘ati” [Mahmudov N, Lutfullayeva D, Odilov Y. 2023] kabilar iboraviy sinonimlarni tizimli tarzda to‘plash va ularni darajalashda muhim manba sifatida xizmat qiladi. Ushbu maqolada aynan iboraviy sinonimlarning darajalanish masalasi, ularning kontekstual ahamiyati va tildagi funksional yuklamasi tahlil etiladi.

Asosiy qism

Iboraviy sinonimlar tushunchasi va ularning o‘rni. Iboraviy sinonimlar — bu ma’nodosh frazeologik birliklar bo‘lib, ular bir xil yoki yaqin ma’noni ifodalashga xizmat qiladi. Ular tilning ifoda imkoniyatlarini kengaytiradi, uslubiy boylikni ta’minlaydi va nutqning emotsional ta’sirchanligini oshiradi. Masalan, “ko‘z oldidan g‘oyib bo‘lmoq” iborasining sinonimlari sifatida “ko‘zdan yo‘qolmoq”, “ko‘zdan g‘oyib bo‘lmoq”, “yo‘qolib qolmoq” kabi iboralarni keltirish mumkin. Bu iboralar bir-biriga yaqin ma’noni ifodalasa-da, ularning stilistik yuklamasi va qo‘llanilish konteksti farq qiladi.

Sinonimik darajalanish tushunchasi . Sinonimik darajalanish — bu ma’nodosh iboralarni semantik (ma’no), stilistik (uslubiy), ekspressiv (emotsional) mezonlar asosida tartiblab, ularning qanday farq qilishini aniqlash jarayonidir. Bunday darajalanish tilshunoslikda iboralarning ma’no yaqinligi asosida farqlanishiga, ularning kuchli yoki zaif sinonimlik darajasini belgilashga imkon beradi.

Sinonimik darajalanish quyidagilarga asoslanadi:

1. Ma’no darajasi – iboralar orasidagi semantik yaqinlik.
2. Stilistik daraja – iboraning qanday uslubda ishlatilishi (rasmiy, norasmiy, badiiy, og‘zaki).
3. Ekspressiv-emotsional yuklama – ibora orqali berilgan emotsional ifoda kuchi.

Misol sifatida, “*qattiq ishlamoq*” iborasining quyidagi sinonimlarini ko‘rib chiqamiz:

- “jonini jabborga berib ishlamoq”
- “qul-elni unutib ishlamoq”
- “bor kuchi bilan ishlamoq”
- “tinmay ishlamoq”

Bu iboralar har xil uslubda va turli kuchdagi emotsional ifodada bo‘lishi mumkin. Quyidagi jadval bu farqlarni ko‘rsatadi:(Qarang:1-jadval)

1-jadval. Sinonimik darajalanish

Iboralar	Ma’nodoshlik darajasi	Stilistik yuklama	Emotsional yuklama	Qo‘llanilish konteksti
Qattiq ishlamoq	Neytral	Kundalik, rasmiy	Past	Har qanday kontekst
Jonini jabborga berib ishlamoq	Kuchli sinonim	Badiiy, emotsional	Juda yuqori	Badiiy matnlar, ritorika
Qul-elni unutib ishlamoq	Kuchli sinonim	Badiiy	Yuqori	Badiiy va ta’sirchan nutqlar
Bor kuchi bilan ishlamoq	O‘rtacha sinonim	Neytral, rasmiynroq	O‘rtacha	Rasmiy matnlarda, maqolalarda
Tinmay ishlamoq	Zaif sinonim	Kundalik	Past	Oddiy kundalik nutqda

- “*Qattiq ishlamoq*” — asosiy neytral ibora bo‘lib, deyarli barcha uslubda ishlatiladi.
- “*Jonini jabborga berib ishlamoq*” — kuchli sinonim, unda fidoyilik va o‘zini fido qilish obrazli tarzda ifodalangan, badiiy va hissiy ranglar kuchli.
- “*Qul-elni unutib ishlamoq*” — mazmunan chuqurroq bo‘lib, ishlovchining atrofdagi narsalarni unutgan holdagi holatini ifodalaydi.
- “*Bor kuchi bilan ishlamoq*” — bu ibora kuchli mehnatni bildiradi, biroq undan yuqori darajadagilar qadar emotsional yuklama bermaydi.
- “*Tinmay ishlamoq*” — oddiy va tez-tez ishlatiladigan ibora, emotsional ifodasi kuchsizroq.

Sinonimik darajalanish iboralarni faqat ma’nodosh deb ko‘rmasdan, ularning ohangini, uslubiy rang-barangligini, va ta’sirchanlik darajasini aniqlashda muhim ahamiyatga ega. Bu yondashuv orqali biz biror iboraning o‘rniga qaysi alternativ ifoda ko‘proq mos kelishini aniqlashimiz, matnning uslubiga mos tarzda iboralarni tanlashimiz mumkin bo‘ladi.

Iboraviy sinonimlarning turlari

a) **Semantik sinonimlar**

Bu turdagi iboralar bir xil yoki juda yaqin ma’noni ifodalaydi. Masalan, “jonini jabborga berib ishlamoq” va “qul-elni unutib ishlamoq” iboralari har ikkisi ham mehnatga bo‘lgan fidoyilikni ifodalaydi.

b) **Stilistik sinonimlar**

Bu iboralar ma’nodosh bo‘lsa-da, uslubiy jihatdan farqlanadi. Masalan, “ko‘z oldidan g‘oyib bo‘lmoq” iborasi badiiy uslubda qo‘llanilsa, “yo‘qolib qolmoq” iborasi kundalik nutqda ishlatiladi.

c) **Ekspressiv-emotsional sinonimlar**

Bu turdagi iboralar ma’nodosh bo‘lishi bilan birga, kuchli emotsional yuklamaga ega. Masalan, “yuragi orqaga tortmoq” iborasi qo‘rquv yoki xavotirni ifodalaydi, bu iboraning sinonimi sifatida “dili g‘ash bo‘lmoq” iborasini keltirish mumkin.

Quyida iboraviy sinonimlarning darajalanishiga oid amaliy misollar keltiriladi:

- “*Ko‘z oldidan g‘oyib bo‘lmoq*” — bu ibora badiiy uslubda ishlatiladi va obrazli ifodaga ega.
- “*Ko‘zdan yo‘qolmoq*” — bu ibora neytral uslubda qo‘llaniladi.
- “*Yo‘qolib qolmoq*” — bu ibora kundalik nutqda ishlatiladi va sodda ifodaga ega.

Bu misollar iboraviy sinonimlarning uslubiy va semantik darajalanishini ko‘rsatadi.

Iboraviy sinonimlarning stilistik va semantik xususiyatlari. Iboraviy sinonimlar stilistik jihatdan rasmiy, norasmiy, badiiy, ilmiy kabi uslublarda farqlanadi. Semantik jihatdan esa, ular ma’no nozikliklari, emotsional yuklamasi va qo‘llanilish kontekstiga ko‘ra farqlanadi. Masalan, “yuragi orqaga tortmoq” iborasi qo‘rquvni ifodalasa, “dili g‘ash bo‘lmoq” iborasi xavotirni ifodalaydi.

Iboraviy sinonimlarning lug‘atlarda aks etishi. O‘zbek tilida iboraviy sinonimlar “O‘zbek tilining frazeologik lug‘ati”da keng yoritilgan. Bu lug‘atda iboraviy sinonimlarning ma’nolari, qo‘llanilish konteksti va stilistik xususiyatlari ko‘rsatilgan. Shuningdek, “O‘zbek tili sinonimlarining katta izohli lug‘ati”da ham iboraviy sinonimlar haqida ma’lumotlar mavjud.

Xulosa

O‘zbek tilidagi iboralarning sinonimik darajalanishi tilshunoslikda muhim o‘rin tutadi. Sinonim iboralar o‘rtasidagi semantik, stilistik va ekspressiv-emotsional farqlarni aniqlash orqali til vositalaridan aniq va samarali foydalanish imkoniyati yaratiladi. Har bir iboraning o‘ziga xos kontekstdagi ma’no yuklamasi borligi, ularning bir-biriga bo‘lgan yaqinlik darajasini ilmiy mezonlar asosida baholashni taqozo etadi.

Darajalanish yordamida tilshunoslar, adabiyotshunoslar hamda tarjimonlar bir ma’noni ifodalovchi turli iboralardan zarur bo‘lganini to‘g‘ri tanlash imkoniga ega bo‘ladilar. Shuningdek, sinonimik iboralarning darajalanishi nutq boyligini oshirish, matnning uslubiy rang-barangligini ta’minlash hamda ifoda aniqligini kuchaytirish vositasi sifatida ham xizmat qiladi.

Tahlillardan ko‘rinib turibdiki, har bir ibora o‘zining emotsional kuchi, qo‘llanilish sohasi va uslubiy xususiyatlariga ko‘ra boshqa iboradan farqlanadi. Masalan, “qattiq ishlamoq” iborasi

oddiy va neytral bo'lsa, "jonini jabborga berib ishlamoq" badiiy va hissiy jihatdan kuchliroqdir. Bu farqlarni tushunish va to'g'ri tahlil qilish til madaniyatini rivojlantirishda muhim omil hisoblanadi.

Umuman olganda, sinonimik darajalanish tushunchasini chuqur o'rganish o'zbek tilining boy ifoda imkoniyatlarini ochib berishda, tarjimalar sifati va matnlar stilistikasini yaxshilashda muhim ilmiy-amaliy ahamiyat kasb etadi. Shu sababli, mazkur yo'nalishda olib borilayotgan izlanishlar tilshunoslikning nazariy va amaliy rivojiga katta hissa qo'shadi.

Foydalanilgan adabiyotlar ro'yhati:

1. Rahmatullayev Sh, Mahmudov N, Xolmonova Z. O'zbek tili frazeologik lug'ati. T.: G'afur G'ulom nomidagi nashriyotmatbaa ijodiy uyi, 2022.
2. Qobulova G. Frazeologik birliklarning lingvistik va semantic xususiyatlari. O'zbekiston Milliy universiteti xabarlari. 2024.
3. Ganiyeva Sh.A. "O'zbek frazeologizmlari strukturasi" (shakliy va mazmuniy modellashtirish) filologiya fanlari bo'yicha falsafa doktori (Doctor of Philosophy) darajasini olish uchun tayyorlangan dissertatsiya.2017
4. Madrahimov N. "O'zbek tilining frazeologik lug'ati" .2006
5. Mahmudov N, Lutfullayeva D, Odilov Y. O'zbek tili sinonimlarining izohli lug'ati. 2023

РАЗВИТИЕ ТВОРЧЕСКИХ СПОСОБНОСТЕЙ СТУДЕНТОВ НА ЗАНЯТИЯХ ПО РУССКОМУ ЯЗЫКУ И ЛИТЕРАТУРЕ В ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЯХ

Sharipova Sitora Khaydarovna

Teacher at Bukhara University ZARMED,

Doctoral student at Bukhara State University

Email: s.x.sharipova@buxdu.uz

Аннотация: В статье рассматривается важность использования технологии проблемного обучения для развития творческих способностей студентов на уроках русского языка в высших учебных заведениях, современные виды методов, используемых для развития интеллектуальных особенностей, таких как: широкое всестороннее, последовательное и убедительное выражение собственной точки зрения, а также влияние творческого подхода на личностное и профессиональное развитие студента. Рассматриваются особенности преподавания языка, его потенциал как инструмента творческого самовыражения, а также конкретные методики, способствующие развитию воображения, нестандартного мышления и самостоятельности студентов.

Ключевые слова: проблемное обучение, самостоятельная деятельность, проблемная ситуация, дидактический процесс, творческое мышление, креативность, игровые технологии.

Abstract: The article discusses the importance of using problem-based learning technology to develop students' creative abilities in Russian language lessons at higher education institutions, modern types of methods used to develop abilities, such as: broad, comprehensive, consistent and convincing expression of one's own point of view, as well as the influence of a creative approach on the personal and professional development of a student. The article discusses the features of teaching a language, its potential as a tool for creative self-expression, as well as specific methods that contribute to the development of imagination, creative thinking and independence of students.

Keywords: problem-based learning, independent activity, problem situation, didactic process, creative thinking, creativity, gaming technologies.

Креативность — Это не только способность творчески подходить к решению проблем и создавать что-то. Это умение передать свои идеи другим людям, открытость ко всему новому, готовность принимать обстоятельства, один из важнейших навыков XXI века, без которого, невозможно успешное развитие личности в условиях быстро меняющегося мира. Кроме того, к этой группе навыков относят умение адекватно осознавать собственные возможности, понимать свои ошибки и улучшать деятельность.

Современное образование должно не только передавать знания, но и формировать у студентов гибкость мышления, способность к самовыражению и поиску нестандартных решений. Использование новых образовательных методик, интерактивных технологий и творческих заданий позволит эффективно развивать креативность, делая обучение более интересным и продуктивным.

Креативность — это способность человека к нестандартному мышлению, поиску новых идей и оригинальных решений. В образовательном процессе креативность проявляется через:

1. умение анализировать и переосмысливать информацию,
2. способность генерировать новые идеи,
3. самостоятельность в принятии решений,
4. гибкость мышления и адаптивность к новым условиям.

В современном обществе традиционная передача знаний (механическое запоминание информации) постепенно уступает место обучению, основанному на исследовательском и творческом подходе. Креативность играет ключевую роль в развитии:

1. Критического мышления – способность анализировать информацию, находить причинно-следственные связи и избегать шаблонного мышления.
2. Самостоятельности – обучение тому, как находить решения без четких инструкций.
3. Коммуникации – развитие способности ясно и убедительно выражать мысли.
4. Адаптивности – умения быстро реагировать на изменения и искать нестандартные пути решения задач.

Современные образовательные методики все чаще отходят от традиционных лекций в сторону интерактивных форм обучения:

1. Метод кейсов – рассмотрение реальных или вымышленных ситуаций с последующим поиском нестандартных решений.
2. Мозговой штурм – активный обмен идеями в группе.
3. Ролевые игры и дебаты – развитие способности мыслить гибко и аргументированно излагать свою точку зрения.

Когда студенту дается возможность выбирать, каким способом он хочет выразить свои знания (например, написать эссе, создать видеоролик, нарисовать инфо графику), это стимулирует его творческого подхода. Творческое мышление это и есть креативность.

В будущем роль креативности в образовании будет только возрастать. Развитие технологий, искусственного интеллекта и новых форм коммуникации потребует от нас способности к инновациям и гибкому мышлению. Именно поэтому, система образования должна адаптироваться к этим изменениям и активно внедрять творческие методы обучения.

Русский язык как учебная дисциплина играет важную роль в развитии творческого мышления, поскольку сочетает лингвистический анализ, литературное творчество и речевую деятельность. Изучение выразительных средств языка, работа с текстами, написание сочинений и эссе формируют умение мыслить образно, находить нестандартные решения и выражать собственные идеи в оригинальной форме. Анализ литературных произведений, интерпретация художественных текстов и создание собственных текстов способствуют развитию ассоциативного мышления, логики и коммуникативных навыков. Таким образом, русский язык не только обучает грамотности, но и развивает креативность, необходимую в современных профессиях и жизни.

Основная цель данной статьи — проанализировать значение креативности в образовательном процессе и определить эффективные методы её развития на занятиях русского языка. В нашем исследовании рассматриваются вопросы и особенности преподавания языка, его потенциальная роль как инструмента творческого самовыражения, а также конкретные методики, способствующие развитию воображения, нестандартного мышления и самостоятельности студентов.

Развитие креативности на занятиях русского языка включает в себя:

1. освоения выразительных средств языка;
2. анализа и интерпретацию литературных произведений;
3. создания оригинальных текстов (эссе, сочинений, рецензий, сценариев);
4. применения интерактивных и игровых методов обучения;
- 5.использования современных цифровых технологий для развития творческого подхода к обучению;

В данной статье приводятся примеры практических заданий, которые помогают студентам раскрыть свой творческий потенциал, учат находить неожиданные смысловые связи, формулировать оригинальные мысли и выражать их в устной и письменной речи.

Важным аспектом является также роль преподавателя, который должен создавать условия для творческого поиска, мотивировать студентов к нестандартному мышлению, развивать их способность к анализу и аргументации, а также вовлекать в проектную и исследовательскую деятельность.

Таким образом, целью данного исследования является выявление эффективных стратегических подходов, которые позволят преобразовать изучение русского языка и процесс формирования и развития творческого мышления студентов, необходимого для успешного обучения и будущего профессионального развития.

Влияние креативности на личностное и профессиональное развитие студента.

Креативность играет ключевую роль в личном и профессиональном развитии студента, формируя гибкость мышления, способность находить нестандартные решения и адаптироваться к новым условиям. В личностном плане она способствует самовыражению, развитию воображения и уверенности в своих идеях. В профессиональной сфере креативное мышление помогает решать сложные задачи, находить инновационные подходы и эффективно взаимодействовать в команде. Современные работодатели ценят специалистов, способных мыслить нестандартно и предлагать новые идеи, поэтому развитие креативности дает студенту конкурентное преимущество и открывает больше новых возможностей для успешной карьеры.

Особенности преподавания русского языка, способствующие развитию креативности. Преподавание русского языка обладает рядом особенностей, способствующих развитию креативности у студентов. Во-первых, язык сам по себе является инструментом самовыражения, а работа с текстами, анализ художественных произведений и создание собственных сочинений стимулируют воображение и образное мышление. Во-вторых, изучение выразительных средств (метафор, эпитетов, сравнений) учит студентов находить нестандартные языковые решения и формулировать оригинальные образные мысли. В-третьих, интерактивные методы, такие как ролевые игры, дискуссии, написание текстов в разных жанрах (эссе, сценарии, интервью), развивают ассоциативное мышление и способность к импровизации. Также важную роль играет проектная деятельность, например, создание блогов, литературных журналов или видеороликов, что помогает студентам применять языковые навыки в творческом формате. Таким образом, преподавание русского языка не только формирует грамотность, но и развивает гибкость мышления, креативный подход к решению задач и навыки эффективной коммуникации.

Методы и примеры развития креативности на занятиях по русскому языку.

Развитие креативности на уроках русского языка возможно через применение различных методов и приемов, которые помогают студентам мыслить нестандартно, находить оригинальные решения и выражать свои идеи в творческой форме. Среди наиболее эффективных методов можно выделить игровые технологии, работу с текстами, развитие ассоциативного мышления и использование мультимедийных технологий.

Игровые технологии:

Игровые методы делают процесс обучения увлекательным и способствуют развитию воображения, импровизации и нестандартного мышления. Игры помогают снять напряжение, повысить интерес к предмету и вовлекают студентов в активное взаимодействие с материалом.

Примеры игровых заданий:

1. «Живая грамматика» – студенты распределяют между собой роли частей речи и создают мини-спектакль, где каждый персонаж взаимодействует в соответствии со своими грамматическими особенностями.

2. «Лингвистический детектив» – преподаватель предлагает текст с ошибками или стилистическими неточностями, а студенты должны выступить в роли «детективов» и выявить их.

3.«Словесный калейдоскоп» – учитель создает условия для развития интеллектуальных способностей учащихся через игровые упражнения со словами, пословицами, крылатыми выражениями путем комбинирования слов по определенному правилу, расшифровки, отгадывания и превращения слов. Участники могут придумывают необычные словосочетания, объясняя их смысл.

Игровые технологии не только улучшают восприятие материала, но и помогают студентам развивать ассоциативное и креативное мышление.

Анализ текста:

Анализ и создание текстов – важный способ развития креативности, поскольку он требует не только понимания языка, но и способности интерпретировать, перерабатывать информацию и выражать мысли в оригинальной форме.

Примеры заданий:

1.«Альтернативные финалы» – студенты предлагают новые концовки известных литературных произведений, учитывая стиль и характер персонажей.

2.«Переписывание истории» – создание художественного текста от лица второстепенного персонажа или изменение перспектив повествования.

3. «Мини-рассказ по ключевым словам» – преподаватель предлагает набор слов, из которых студенты должны составить связной и логично-правильный, стилистически соответствующий рассказ.

Работа с текстами помогает развивать логику, воображение, образного мышления и способность к анализу, а также улучшает навыки письменной речи.

Развитие креативности на занятиях русского языка требует применения разнообразных методик, ориентированных на активное взаимодействие студентов с языком и текстом. Игровые технологии, работа с художественными произведениями, упражнения на ассоциации и современные цифровые инструменты помогают формировать гибкость мышления, творческого подхода и навыков самовыражения. Такой подход делает обучение не только эффективным, но и интересным, мотивируя студентов к активному изучению русского языка и развитию их интеллектуального потенциала.

Практический опыт и примеры успешных методик.

Развитие креативности на занятиях русского языка подтверждается практическим опытом и навыком педагогов, использующих нестандартные методы обучения. Многие преподаватели отмечают, что применение игровых технологий, проектных заданий и цифровых инструментов повышает мотивацию студентов, делает процесс обучения более увлекательным и эффективным. Ниже приведены реальные кейсы из образовательной практики, демонстрирующие успешные методики развития творческого мышления.

Кейс 1: Литературный блог как способ вовлечения студентов.

Описание: В университете БухГУ преподаватель предложил студентам создать литературный блог, где они могли публиковать рецензии на книги, эссе о языковых явлениях или творческие работы. Например, найдите тексты лирических стихотворений в которых времена года изображены по-разному. Выпишите фрагменты, сопоставьте их. Сделайте выводы какие средства художественной выразительности авторы чаще всего использовали для описания времен природы. Сопоставьте времен года и в искусстве. Как изображается времена года в искусстве. Проект длился на один семестр, а в конце был проведен конкурс на «Лучшее исследование».

Результаты:

- 80% студентов отметили, что стали более внимательно относиться к языковым средствам выразительности и стилю речи.

1. Улучшились навыки аргументации в устной и письменной речи.

2. Учащиеся научились анализировать литературные произведения в доступной объективной форме.

Отзыв студента:

“Этот проект помог мне не только улучшить письменные навыки, но и преодолеть страх перед публичным выступлением мнений. Теперь я знаю, на какие качества текста надо обращать внимание для того, чтобы написать лирические произведения, сделать его интересным для читателя.”

Кейс 2: Театрализация литературных произведений

Описание: В одной группе преподаватель предложил студентам инсценировать отрывки из произведений русской классики, но с измененными жанровыми стилями (например, «Преступление и наказание» в жанре комедии, «Горе от ума» в стиле детектива).

Результаты:

1. Студенты глубже поняли содержание произведений и их скрытые смыслы и мотивы.

2. Повысился уровень интереса к литературе.

3. Учащиеся проявили креативные подходы, работая с текстами и изучая их глубже.

Отзыв студента:

“Мы не просто заучивали текст, а создавали собственные интерпретации. Это было весело и познавательно!”

Кейс 3: Интерактивные викторины и игры на развитие ассоциативного мышления.

Описание: В университете преподаватель использовал платформы Kahoot и Quizizz для проведения интерактивных викторин. Также на занятиях применялся метод «случайных ассоциаций»: студенты получали набор несвязанных слов и должны были придумать связной рассказ.

Результаты:

1. Повысилась активность студентов в аудитории.

2. Улучшилась способность к быстрому анализу и нестандартному мышлению.

3. Учащиеся стали лучше запоминать сложные лексические и грамматические конструкции.

Отзыв студента:

“Викторины и игры сделали занятия живыми. Благодаря таким приемам, я стала запоминать новые слова быстрее и научилась мыслить нестандартно”.

Выводы об эффективности предложенных методов

Анализ практических кейсов показывает, что креативные методики способствуют не только повышению интереса студентов к изучению русского языка, но и развитию ключевых навыков, таких как критическое мышление, способность к импровизации и самостоятельному анализу информации. Использование игровых технологий, проектной деятельности и мультимедийных инструментов делает обучение более живым и продуктивным, а отзывы студентов подтверждают их эффективность и результативность.

Заключение

Развитие креативности в обучении играет ключевую роль в формировании гибкого и нестандартного мышления у студентов, что особенно важно в современном мире, требующем инновационных решений и способности адаптироваться к новым условиям. Занятия по русскому языку, благодаря своей многогранности, становятся эффективной площадкой для развития творческих способностей. Применение игровых технологий, работы с текстами, мультимедийных инструментов и ассоциативных методик помогает студентам научиться анализировать информацию, выражать свои мысли оригинально и уверенно, а также находить нестандартные подходы к решению задач.

Роль преподавателя в этом процессе сложно переоценить. Он не только передает знания, но и создает условия для творческого поиска, мотивирует студентов к самостоятельной работе, поддерживает их инициативу и помогает раскрыть потенциал. Важно не просто дать теоретическую базу, но и научить студентов использовать язык как инструмент самовыражения и критического осмысления действительности. Процессы

цифровизации, несмотря на определенные опасения, лишь подчеркивают высокий статус педагогического труда, важность личного контакта учителя с учащимися. В настоящий момент в рамках реализации образования ведется широкомасштабная работа по модернизации отечественной системы образования, в том числе обновления материально-технической базы образовательных организаций всех уровней. Особое внимание уделяется и подготовке педагогических кадров, привлечению и удержанию в профессии молодых специалистов. Очевидно, что эти и другие меры, предпринимаемые в последние годы, дадут свои плоды, и прежде всего в качестве подготовки наших учащихся.

Поэтому перспективы развития креативных методик в преподавании русского языка связаны с дальнейшей интеграцией интерактивных технологий, разработкой новых игровых форм обучения, применением цифровых инструментов и развитием проектной деятельности. Использование современных мультимедийных платформ, искусственного интеллекта и геймификации может сделать изучение не только русского языка, но и другие языки более увлекательным и продуктивным.

По мнению преподавателей наших вузов, основными приоритетными методами преподавания теоретических дисциплин являются обучающие информационно-компьютерные программы, проблемно-ориентированное, дистанционное обучение, интерактивный метод обучения, деловые игры с демонстрацией каждого этапа изучения наглядными материалами. Причем, большая часть профессорско-преподавательского состава подтверждает применение указанных современных, высокоинформативных методов обучения при преподавании различных дисциплин в учебном процессе.

Таким образом, можно сделать вывод, что данные методики обучения не только повышают уровень подготовки студентов, но и способствуют их всестороннему интеллектуальному развитию, формируя у них успешную профессиональную и личностную основу.

Использованная литература:

1. Лингводидактика и методика обучения иностранным языкам [Электронный ресурс] : учеб.-метод. пособие / авт.-сост. О. В. Миха-лева ; Владим. гос. ун-т им. А. Г. и Н. Г. Столетовых. – Владимир: Изд-во ВлГУ, 2022. – 136 с.
2. Xaydarovna, Sharipova Sitora. "Ijtimoiy-Gumanitar Fanlarni O'qitishda Muammoli Ta'lim Texnologiyasidan Foydalanish Mazmuni." Journal of Innovation in Education and Social Research 2.5 (2024): 24-29.
3. Xaydarovna S. S. Bo'lajak Chet Tili O'qituvchilarning Lingvodidaktik Kompetentsiyasi: Shakllanish Darajasini Aniqlash Usuli //Journal of Research in Innovative Teaching and Inclusive Learning. – 2024. – T. 2. – №. 4. – С. 93-98.
4. Шарипова С. OLIY TA'LIM MUASSASALARIDA RUS TILI FANINI O'QITISHDA MUAMMOLI TA'LIMDAN FOYDALANISH //Ижтимоий-гуманитар фанларнинг долзарб муаммолари/Актуальные проблемы социально-гуманитарных наук/Actual Problems of Humanities and Social Sciences. – 2023. – T. 3. – №. 6. – С. 302-307.
5. Sharipova S. K. The essence of using problem-based learning technologies in teaching physical education disciplines //BIO Web of Conferences. – EDP Sciences, 2024. – T. 120. – С. 01056.
6. Sharipova S. K. LANGUAGE EDUCATION AS A RESULT OF LEARNING THE LANGUAGE AND CULTURE OF THE PEOPLE //GOLDEN BRAIN. – 2024. – T. 2. – №. 22. – С. 120-133.
7. Sharipova S. K. LANGUAGE EDUCATION AS A RESULT OF LEARNING THE LANGUAGE AND CULTURE OF THE PEOPLE //GOLDEN BRAIN. – 2024. – T. 2. – №. 22. – С. 120-133.
8. Sharipova S. H. Technology of Problem Learning as a Tool for the Development of Students' Independent Work //Miasto Przyszłości. – 2022. – T. 30. – С. 129-130.

TEZAURUS YARATISHDA MUNOSABAT VA ALOQA TUSHUNCHASI

Yo'ldoshaliyeva Muyassar Yo'ldashboy qizi

O'zMU Kompyuter lingvistikasi yo'nalishi

1- bosqich magistranti

E-mail: muyassarjdi@gmail.com

ORCID-0009-0003-3931-5720

Annotatsiya: Ushbu maqolada tezaurus lug'atlari, ular o'rtasidagi assotsiativ munosabat, ierarxik munosabat ularning bir-biridan farqi, bu tushunchalarning mazmun mohiyati ochib berilgan.

Kalit so'zlar: korpus, tezaurus, lug'at, assotsiativ munosabat, ierarxik munosabat.

Abstract: In this article, the content of thesaurus dictionaries, the associative relationship between them, the hierarchical relationship and their difference from each other are revealed.

Keywords: corpus, thesaurus, glossary, associative relation, hierarchical relation.

Аннотация: В данной статье раскрывается содержание тезаурусных словарей, ассоциативная связь между ними, иерархическая связь и их отличие друг от друга.

Ключевые слова: корпус, тезаурус, словарь, ассоциативной отношение, иерархическое отношение.

XX asrning so'nggi o'n yilliklaridanoq o'zbek tilining qo'llanish doirasi nihoyatda kengaydi. 1989-yil 21-oktyabrda "O'zbek tiliga davlat tili" maqomining berilishi ilk qadamlardan biri hisoblanadi. Bu esa tilimizning rivoji uchun keng imkoniyatlarni hozirladi. Uning isboti sifatida ilmiy asosda rivojlantirishga qaratilgan tadqiqotlar, tilimizning o'ziga xos xususiyatlariga bag'ishlangan ilmiy, ommobop kitoblar, o'quv qo'llanmalar, ilmiy jurnallar va yangidan yangi ilmiy lug'atlarni aytish o'rnlidir. Ammo minglab yillar davomida rivojlangan inson tafakkuri uchun bugungi kunda birlamchi manbaa bo'lish, ya'ni axborot olish va berish yagona manbaga aylandi. XX asrning buyuk kashfiyoti kompyuterning rivojlanishi bevosita tilning rivoji uchun ham ta'sir o'tkazdi. Natijada, tilning kompyuter, internet tiliga aylanishiga olib keldi. Yangi yo'nalishlar matematik lingvistika, kompyuter lingvistika, korpus lingvistikasi kabi sohalarining shakllanishiga zamin hozirladi.

XXI asr axborot texnologiyalari asrida zamonning shiddat bilan rivojlanishi kompyuter lingvistikasini rivojlanishiga ehtiyoj sezmoqda. Aynan mana shu yo'nalishgina o'zbek tilining jahon miqyosiga olib chiqadi va dunyo tillari orasida o'z o'rnini egallashiga imkon beradi. O'zbek tili korpus (www.uzbekcorpus.uz) yaratilishi, uning yillar davomida boyib borishi, tilning jahonda o'z o'rnini topishiga turtki bo'ladi. Korpusning boyishi tarjima imkoniyatini kengaytiradi. Til o'z-o'zidan elektronlashadi. Bugungi kunda xorijiy tildagi birorta kitob yoki maqola matnini o'zbek tiliga tarjima qilish uchun *google translated*dan foydalansak, gap semantik va struktur jihatdan xato bo'lishi aniq, korpusning boyishi, o'zbek tilining tezaurusi va ontologiyasi yaratilishi oldimizda turgan mana shu muammoni bartaraf etadi.

"Tezaurus" so'zi yunon tilidan kelib chiqqan bo'lib, o'zbek tilida qo'riqxon, ya'ni xazina degan ma'noni anglatadi. Yanada aniqroq aytsak, ma'lum mavzuga birlashgan terminlardan iborat lug'at. Tezaurusda maxsus sohaning ta'riflari, tasniflari, atamallari, tushunchalari to'liq qamrab olinadi. Tezaurus lug'atga o'xshab ketadi. Tezaurusda so'zlar alifbo tartibida emas, o'zaro ma'no yaqinligi jihatdan tartiblanadi. Lug'atda bo'lsa buni aksi.

Kompyuter leksikografiyasida lug'atlar mashina o'qiy oladigan formatda bo'lishi, qidiruv tizimiga egaligi, indekslanganligi, kategoriyalashtirilganligi, klasterlashtirilganligi, va foydalanuvchiga undan foydalanish imkoniyatining qulayligi va kengligi bilan xarakterlanadi. [Abduraxmonova N.Z., 2021; 39].

O'zbek tilining tezaurusini yaratish uchun WordNet, RuTez kabi avval yaratilgan tezaurus namunalaridan o'rganish maqsadga muvofiq. Lug'at qanday bo'lishidan qat'iy nazar doim ikki qismdan tashkil topadi. Birinchisi-indeksi, ikkinchisi-izohi. Tezaurusni strukturaviy jihatdan chuqur tahlil qilish ahamiyatlidir.

Axborot-qidiruv tezauruslari odatda ma'lum mavzular uchun yaratilganligi sababli, ularning tuzilishi, asosan, "konseptsiya" va "atama" kabi obyektlarga asoslanadi. Bu odatda ma'lum bir bilim yoki faoliyat sohasi konseptsiyasini ko'rsatadigan so'z yoki ibora sifatida tushiniladi.[Lukashevich N.V., 2010; 26].

Tezaurusni tavsiflab kelgan tavsiflovchilari tanlangan mavzuga mos bo'lishi kerak. Tezaurusga kiritilgan har bir deskriptor belgilangan maydonning alohida tushunchasini ifodalab kelishi kerak. Deskriptorlar bir so'zli va ko'p so'zli bo'lishi mumkin. Deskriptor aniq shakllangan bo'lishi va tezaurusdagi nazarda tutilgan ma'nosi o'quvchi uchun tushinarli bo'lishi lozim. Bundan tashqari axborot-qidiruv tezaurusi o'rtasida aloqalar mavjud. Jins-tur, sabab-oqibat, qism-butun, xomashyo-mahsulot kabi aloqalar tezaurusning muhim jihati hisoblanadi. Ushbu aloqalar hamma tezauruslarni birlashtiruvchi omil hisoblanadi.

Deskriptorlar o'rtasida yana ikkita muhim munosabat bor:

1. Ierarxik munosabat.
2. Assotsiativ munosabat.

Aynan Amerika tezaurusi ierarxik munosabat ustiga qurilgan. Ierarxik munosabat bu-pastdan bitta yuqoriga qarab harakatlangan semantik ma'no deyish mumkin. Masalan, guruch va osh; ildiz-daraxt. Osh o'zbeklarning milliy taomlari sirasiga kiradi. Guruch oshning bir elementi ya'ni taom tarkibidagi bir mahsulot hisoblanadi. Ildiz ham xuddi shunday. Daraxtning shoxlari, barglari, guli, mevasi gul kosa bargi, tanasi va ildizi bor. Daraxt bir oilani tashkil etsa, ildiz uning bir a'zosi hisoblanadi. Ba'zi mualliflar tezaurusdagi ierarxik munosabatni qisqartirishni muhokama qilishgan. Shunday qilib, amaliy tajriba shuni ko'rsatadiki, tezaurusning ierarxik chuqurligi ma'lum bir chegarada to'xtashi kerak. Aks holda bu foydalanuvchi uchun noqulaylik tug'dirishi tabiiy. Shu sababli tezaurusda 9 ta ierarxik chegara qo'yilgan. Ierarxik munosabat tarkibiga jins-tur, qism-butun, yuqori-quyi aloqalari kirib ketadi. Keyingi munosabat assotsiativ munosabat hisoblanib, axborot-qidiruv tezaurusini yaratish shuni ko'rsatadiki, bu munosabat ierarxik munosabat singari sodda emas. Uni aniqlash biroz mushkul. Uni bir jumla bilan: "Ierarxik munosabat va sinonimik munosabatlarning bir qismi bo'lmagan munosabatlar bilrligi"- deya ta'kidlash joiz.

Amerikalik mutaxassislar shuni ta'kidlaydiki, deskriptorlar o'rtasidagi assotsiativ munosabatlarni o'rnatishning eng umumiy qoidasi shuni tavsiflaydiki, agar bir atama ishlatilganda boshqa bir atama nazarda tutilgan bo'lsa, bu munosabat ikkita deskriptor o'rtasida o'rnatilishi kerak. Bundan tashqari bir atama boshqa bir atamaning zarur elementi hisoblanadi. Jabra so'zini xayolimizga keltirishimiz bilan baliq ko'z o'ngimizda gavdalanadi.

Ierarxik munosabat singari assotsiativ munosabat tarkibida ham muhim aloqalar mavjud.

1. Harakatlar va ularning natijasi: bichish- ko'ylak
2. Moddalar va ularning xususiyatlari: zahar- zaharlilik
3. Sabab va oqibat bilan bog'liq bo'lgan tushunchalar: o'lim-motam, bola tug'ilishi- aqiqa
4. Tushunchalar va ularning o'lchov birliklari o'rtasidagi munosabatlar: elektr toki- amper
5. Jarayonlar va ularning asboblari: ovchi- kamon
6. O'rganish obyekti yoki mutaxassis: tilshunoslik-tilshunos.

Assotsiativ munosabatga tayangan holda tezaurus yaratish bir qancha noaniqliklarga olib kelishi mumkin. Ingliz tilshunos olim Gerd shuni ta'kidlaydiki: "Bu munosabatdan foydalanishda ma'lum bir mavzu bo'yicha matnlardan foydalanish, bunday matnlarni tahlil qilish bizga ma'lum bir mavzu sohasiga xos bo'lgan semantik munosabatlarning haqiqiy turlarini ochishga imkon beradi. Ushbu yondashuv bilan ma'lumot qidirish tezaurusida aniqlangan assotsiativ munosabatlar matnda mavjud bo'lgan semantik munosabatlarga mos keladi" [Lukashevich N.V.,

Dobrov V., 2000; 15]. Xulosa o‘rnida axborot-qidiruv tezaurusini shakllantirishda birinchi vazifa tezaurusga kiritish uchun munosabatlarni tanlashdir.

Yuqoridagilardan shuni ta’kidlash joizki, o‘zbek tilining korpusi yaratilishi ona tilimizning sof holda kelajak avlodga saqlanib qolishi uchun bir qadam bo‘ldi. Tezauruslarning yaratilishi tarjima imkoniyatlarini kengaytiradi. Tilning boyishi, korpusga kiritilishi o‘zbek tilining jahonda o‘rnini topishiga zamin hozirlaydi.

Foydalanilgan adabiyotlar ro‘yxati:

1. Abduraxmonova N.Z. Kompyuter lingvistikasi (darslik,) - Toshkent: Nodirabegim, 2021. – 394 b.
2. Abduraxmonova N.Z., Gulruhsor, X. (2024). O ‘zbek tili tezaurusini yaratishda standart me’yorlar. «CONTEMPORARY TECHNOLOGIES OF COMPUTATIONAL LINGUISTICS», 2(22.04), 464-469.
3. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. Москва, 2010, 387
4. Лукашевич Н.В., Dobrov B. Thesaurus-Based Structural Thematic Summary in Multilingual Information Systems - Machine Translation Review, - N 11, December 2000. – pp. 10-20. (<http://www.bcs.org.uk/siggroup/nalatran/mtreview/mtr-11/mtr-11-8.htm>). 2000a.
5. Лукашевич Н.В. Dobrov B., Thesaurus as a Tool for Automatic Detection of Lexical Cohesion in Texts. In Proceedings of 5th JADT. 2000. – pp. 155-162. 2000b.
6. www.uzbekcorpus.uz
7. Gruber T.R. A translation approach to portable ontologies. Knowledge Acquisition, 5(2). 1993. – pp. 199-220.
8. Шевченко Н.В. Основы лингвистики текста. – М.: Приор-издат, 2003.
9. Шемакин Ю.И. Тезаурус в автоматизированных системах управления и информации. - М: Военное изд-во министерства обороны СССР, 1974. - 192 с.

QARAQALPAQ HÁM TÚRK TILLERINÍN PARALLEL KORPUSÍN JARATÍW MÁSELESİ

Ótemisov Aziz Zarlıqbaevich

Qaraqalpaq mámleketlik universiteti
Ámeliy filologiya kafedrası baslıǵı,
filologiya ilimleri boyınsha
filosofiya doktori (PhD), docent.
E-mail: utemisov.aziz@mail.ru

Sharbaev Jaras Bayrambay ulı

Qaraqalpaq mámleketlik universiteti
Lingvistika: qaraqalpaq tili qánigeligi magistrantı
E-mail sharbaevj77@gmail.com

Annotaciya: Bul maqalada qaraqalpaq hám túrk tilleriniń parallel korpusın jaratıwda orınlanıwı kerek bolǵan wazıypalar haqqında sóz etiledi. Maqalada tekstlerdi tayarlaw, tayın tekstler kompleksinen paydalanıw, olardaǵı abzac, gáp, sóz hám sóz dizbeklerin sáykeslestiriw basqıshları, metodlar, programmalar, programmalıq támiynatlar haqqında maǵlıwmatlar berilgen.

Gilt sózler: Korpus lingvistikası, korpus, parallel korpus, tekstlerdi sáykeslestiriw, tokenizaciya, stemmatizaciya, Lingtrain Alignment studio, Humaling, Abbyy Aligner, Trados, Winaling, Wordfast tools, Giza++, AntConc.

Abstract: In this article, we will talk about the tasks that must be performed when creating a parallel corpus of the Karakalpak and Turkic languages. The article provides information on the preparation of texts, the formation of a set of ready-made texts, the stages of matching paragraphs, sentences, words and phrases, methods, programs.

Keywords: Linguistic corpus, corpus, parallel corpus, text matching, tokenization, stemmatization, Lingtrain Alignment studio, Humaling, Abbyy Aligner, Trados, Winaling, Wordfast tools, Giza++, AntConc.

Аннотация: В данной статье речь пойдет о задачах, которые необходимо выполнить при создании параллельного корпуса Каракалпакского и турецкого языков. В статье даны сведения о подготовке текстов, формировании из комплекса готовых текстов, этапах согласования в них абзаца, предложения, слова и словосочетания, методах, программах.

Ключевые слова: Лингвистический корпус, корпус, параллельный корпус, согласование текстов, токенизация, стемматизация, Lingtrain Alignment studio, Humaling, Abbyy Aligner, Trados, Winaling, Wordfast tools, Giza++, AntConc.

Kirish. XXI ásirniń global mashqalalarınan biri tábiyiy tillerdiń milliy ózgesheligin saqlap qalıwdan ibarat. Dúnya tilleriniń elektron korpusların jaratıw hám rawajlandırıwda NLP hám de til texnologiyalarına baylanıslı izertlewlerdi izshil tárizde alıp barıw áhmiyetli wazıypaǵa aylandı. [Abduraxmonova N., 2021: 18]. Ásirese, búgingi qaraqalpaq tiliniń keleshek áwladlarǵa tolıq jetip barıwı, rawajlanıwı, onıń saqlanıwı, qaraqalpaq tiliniń jáhán kólemindegi abıroyınıń artıwı, onıń ilim-pán tili bolıwına dáslepki adım tildiń korpusın jaratıw bolıp tabıladı. Biz bul maqalamızda tuwısqan túrkiy tiller bolǵan qaraqalpaq hám túrk tilleriniń parallel korpusın jaratıw boyınsha eń dáslep orınlanıwı kerek bolǵan wazıypalar haqqında sóz etiwdi maqul kórdik.

Parallel korpus – bul awdarma qılınǵan tekstlerdiń juplıǵı bolıp tabıladı. Tiykarınan parallel korpuslar tek ǵana eki tilde jaratıladı.

Parallel korpusıǵı tekstlerdiń awdarma baǵdarı boyınsha túrleri tómendegilerden ibarat:

- **bir baǵdarlı** (bir tildegi tekst ekinshi tilge awdarmalanǵan);

- **eki tárepleme** (bir tildegi tekst ekinshi tilge awdarmalanad hám kerisinshe);
- **kóp baǵdarlı** (bir tekst kóp tilge awdarmalanadı). [M.Abjalova., 2022: 41-45].

Parallel korpuslar da óz ishinde haqıyqıy parallel korpus hám salıstırmalı korpuslar bolıp bólinedi. Biz maqalamızda sóz etetuǵın másele tek ǵana haqıyqıy parallel korpusqa baylanıslı.

Búgingi kúnde qaraqalpaq xalqınıń tanımalı jazıwshı hám shayırlarınıń, folklorlıq shıǵarmalarınń túrk tiline awdarılǵan nusqaların maqtanış penen tilge alsaq arızıydı. Ásirese, Kayhan Yúkseler tárepinen túrk tiline awdarma islengen T.Qayıpbergenovtıń “Qaraqalpaq qızı” romanı, folklorlıq shıǵarmalardan Djeyhun Vedat Öztürk tárepinen awdarılǵan “Qırq qız” dástanınıń túrk tilindegi awdarmaları bar. Bulardan basqa da túrk tiline awdarma islengen shıǵarmalar bar. Máselen, Berdaq hám Ájiniyaz shayırlardıń kóplegen qosıqları, folklorlıq shıǵarmalardan “Edige” dástanınıń birqansha awdarma variantlarınan da paydalanıwǵa boladı. Bunday materiallardıń barlıǵı eki tilli kishi kólemli parallel korpuslardı jaratıwǵa tiykar bolıp xızmet etedi. Keyin korpusıtaǵı materiallardı tolıqtırıp barıw kerek boladı.

Parallel korpusınıń abzallıqları tómendegilerden ibarat:

1. Tillerdi oqıtıwda tiykarǵa baza xızmetin atqaradı;
2. Eki til ortasındaǵı uqsashlıq hám ayırıqshalıqlardı baqlawǵa imkaniyat jaratadı;
3. Tildegi bolıp atırǵan ózgerislerdi lingvistikalıq tárepten analizlewge imkaniyat jaratadı.

Tiykarǵı bólim. Eń dáslep korpus dúziw jumısları (parallel korpus) eki tildegi tekstlerdi jıynawdan baslanadı. Bunday jumıslarda barqulla mashına awdarması menen qosa sheber awdarmashılardıń jumıs barısında bolıwı yaqı awdarma islengen tayın materiallardan paydalanǵan maqsetke muwapıq. Máselen, awdarma procesinde awdarmashılardıń mıtájlik mashına úlken kólemli tekstte túrli stillik ózgesheliklerdi túsinbewi yaqı qosımtalardıń qáte qosılıw halatlarında anıq seziledi. Bunday jaǵdaylar mashına awdarmasında kóp boladı, jumıstıń mine usı názik táreplerin túsinetuǵın awdarmashınıń bolıwı maqsetke muwapıq. Biraq, awdarmashıǵa qoyılatuǵın talaplar hám sheklewler de joq emes, olardan tekstke dóretiwshilik qatnas jasawǵa jol qoyılmaydı, teksttegi eki-úsh gápti qısquartıp bir gáp formasında beriwge bolmaydı, qanday da bir sebepler menen gáp mazmunın jarıtıp bere almaw sıyaqlılar kiredi. Sózimizdiń basında ayılǵanıday biz parallel korpus jaratıwda T.Qayıpbergenovtıń túrk tiline awdarma islengen “Qaraqalpaq qızı” romanınan hám folklorlıq shıǵarmalardan “Qırq qız” dástanı materiallarınan paydalansaq boladı.

Hárqanday parallel korpusı jaratıw barısında eń áhmiyetli mashqalalardan biri bul tekstlerdi sáykeslestiriw basqıshında boladı. Bunda eki tildegi tekstler teması, ondaǵı gápler, sózler, grammatikalıq qurılısı, dizbekler boyınsha bir-birinen parıqlanadı. Bul mashqalanı avtomatikalıq sheshiwdiń jolın V.Zaxarov, S.Bogdanovlar bılay kórsetedi: *gápler boyınsha sáykeslestiriw, grammatikalıq qurılısı boyınsha sáykeslestiriw, sóz dizbekleri hám sózler boyınsha sáykeslestiriw*. [Захаров В., Богданова С.2020: 61].

Biraq, búgingi kúnde tekstlerdi sáykeslestiriwde sózlerdi sáykeslestiriw metodınan paydalanbaǵan maqul, sebebi, bul usıl hátteki dástúriw awdarma jumıslarında da ózin aqlamaydı hám kútilgen nátiyjeni bermeydi.

A.V.Zubov, I.I.Zubovlar original tekstlerdegi gáplerdi durıslaw hám sáykeslestiriwdiń 6 túrli shamalıq sáykesliklerdiń ushırawın kórsetip ótedi:

- 1) original teksttegi bir gáp áyne bir gáp penen awdarma islenedi;
- 2) eki gáp bir góp penen awdarma islenedi;
- 3) original teksttegi eki gáp bir gápke teńlestirilip awdarma islenedi;
- 4) eki gáp bir gáp penen awdarma islenedi, biraq gáplerdiń ishki shegarası original teksttegi menen awdarma tekstindegi gáplerge sáykes kelmeydi;
- 5) original teksttegi gáp awdarmalanbaydı, sebebi mazmunlıq jaqtan buǵan qájjet qalmaydı;
- 6) awdarma tekstindegi gáp original tekstte bolmawı múmkin. [Зубов А. В., Зубова И. И., 2004: 61].

Mashina awdarmasida tekstleri avtomat sáykeslestiriv procesi morfologiyalıq analiz benen de tıgız baylanışlı esaplanadı. Bunda tokenizaciya, lemmatizaciya, stemmatizaciya sıyaqlı metodlardan paydalanılardı.

Tokenizaciya – tábiyyi tildegi til birlikleri jeke grammatikalıq mánilerge ajıratıp beriledi, sózformalar anıqlanadı.

Lemmatizaciya – sózformaların sózliktegi forması anıqlanadı.

Stemming – sózdiń tiykarı bólege anıqlanadı.

Tekstlerdi sáykeslestirivdiń tómendegi túrleri bar:

- *sózbe-sóz sáykeslestiriv (word alignment)*.
- *gáplerdi sáykeslestiriv (sentence alignment)*.
- *abzac boyınsha sáykeslestiriv (paragraph alignment)*. [M.Abjalova.,2022:41-45].

Tekstlerdi sáykeslestirivshi programmalar da bar bolıp, olardan Humaling, Abbyy Aligner, Trados, Winaling, Wordfast tools, Giza++, AntConc h.t. atap ótiw orınlı. Bunday programmalar gáplerdi belgili uzınlıgı boyınsha salıstırıp, tekstlerdi paraqlarğa bóliw, irkilis belgilerin analiz qılıw sıyaqlı jumslardı ámelge asıradı hám adam-mashina formasında jumıs isleydi. Bulardan basqa GitHub platformasında tekstlerdi sáykeslestirivshi programmalar da bar. Olardan lingtrain_aligner programmalıq kodı az sanlı xalıqlar sóylewshi tillerdegi tekstleri sáykeslestiriv ushın jaratılğan. Tağı da Lingtrain Alignment Studio onlayn programmasınan da paydalanıwğa boladı. Bul web programma bolıp, bir waqıttıń ózinde tekstlerdi sáykeslestiriv, maǵlıwmatlar bazası formasında, .pdf, .doc hám parallel korpus formalarında kirgizilgen maǵlıwmatlardı júklep alıw imkanıyatın beredi. [<https://habr.com/ru/articles/575898/>]

Juwmmaq. Búgingi informaciya-texnologiyalardıń jedel pát penen rawajlanıp atırğan bir dáwirinde ilim-pánniń hesh bir tarawı itibardan shette qalmawı kerek, olardıń qatarında til bilimi de bar. Góneriw aldında turğan, hám kóp waqıttı hám kóp miynetti talap etetuǵın dástúriy metodlarğa salıstırǵanda kóp waqıttı únemlep, unamlı nátiyje beretuǵın sanlı texnologiyalarğa tiykarlangan jańa usıllardan paydalanıw waqtı keldi. Bunday metodlardan paydalanıwda til biliminde korpuslar jaratıwdıń, olardan paydalanıwdıń ornı ayrıqsha. Ásirese tillerdi bir-birine salıstırıp úyreniwde parallel korpuslar oǵada áhmiyetli orındı iyeleydi. Mine usınday máselelerdi sheshiw búgingi korpus lingvistikasınıń áhmiyetli wazıypalarınan biri esaplanadı.

Paydalanılğan ádebiyatlar:

1. Abduraxmonova N. O`zbek tili elektron korpusining kompyuter modellari (monografiya) /Toshkent: Muharrir, 2021, B – 202.
2. Abjalova M. Korpus lingvistikasi (uslubiy qollanma)/ Toshkent: Bookmany Print, 2022, B –103.
3. Захаров В., Богданова С. Корпусная лингвистика: учебник. 3-е изд., перераб. – СПб.: Изд-во С.-Петербур. ун-та, 2020, С – 61.
4. Зубов А. В., Зубова И. И. Информационные технологии в лингвистике: учеб. пос. М.: Издательский центр «Академия», 2004, С – 208.
5. <https://habr.com/ru/articles/575898/>

QORAQOLPOQ TILIDAGI SODDA GAP TURLARINI SINTAKTIK TAHLIL BOSQICHI UCHUN TAYYORLASH

O'temisov Aziz

QDU, Amaliy filologiya kafedrası mudiri, docent

E-mail: utemisov.aziz@mail.ru

Mambetullaeva Biybihajar

Qoraqolpoq Davlat Universiteti 1-kurs magistranti

Annotation: Sentence parsing is one of the most important tasks in natural language processing (NLP). Simple sentences are the most basic unit of language in communication and include the stages of morphological and syntactic analysis. In this article, we attempted to develop the syntactic structure of simple sentences and their types for the syntactic analyzer stage and implement it using Python.

Keywords: NLP, simple sentence, Universal Dependencies, parser, syntax tags.

Аннотация: Синтаксический анализ предложения — одна из важнейших задач обработки естественного языка. Простые предложения являются самой базовой единицей языка в коммуникации и включают в себя этапы морфологического и синтаксического анализа. В данной статье мы попытались разработать синтаксическую структуру простых предложений и их типы для этапа синтаксического анализа и реализовать её на языке Python.

Ключевые слова: NLP, простые предложение, Universal Dependencies, синтаксический анализатор, синтаксические теги.

Annotatsiya: Gapni tahlil qilish tabiiy tilni qayta ishlashning eng muhim vazifalaridan biridir. Oddiy jumlar muloqotda tilning eng asosiy birligi bo'lib, morfologik va sintaktik tahlil bosqichlarini o'z ichiga oladi. Ushbu maqolada biz jumlarning sintaktik tuzilishini va ularning turlarini tahlil qilish bosqichi uchun ishlab chiqishga va uni Python tilida amalga oshirishga harakat qildik.

Kalit so'zlar: NLP, sodda gap, Universal Dependencies, sintaktik analizator, sintaktik teglar.

XX asr o'rtalaridan boshlab raqamli texnologiya jadal rivojlanishi bilan, bir qancha yangi fan sohalari oldiga dolzarb vazifalarni qo'ydi. Ta'kidlash joizki, tabiiy tilni qayta ishlashda matn analizatorlarini ishlab chiqish ustuvor masalalardan biri bo'lib kelmoqda. Chunki kompyuter asri bo'lgan XXI asrda oltindan qimmatroq bo'lgan ma'lumotlarni ziyon yetkazmay qabul qilish, ishonchli saqlash va eng qulay usulda boshqalarga yetkazib berish til bilan chambarchas bog'liq bo'lgan vazifalardan biriga aylandi [Po'latov S, Muhammedova S., 2009: 36].

Matnni tahlil qilish hozirgi kunda korpus lingvistikasida alohida o'rganilmoqda. Til korpusi matnlar to'plami bo'lib, undagi matnlar quyidagicha tahlil qilingan ko'rinishda bo'ladi:

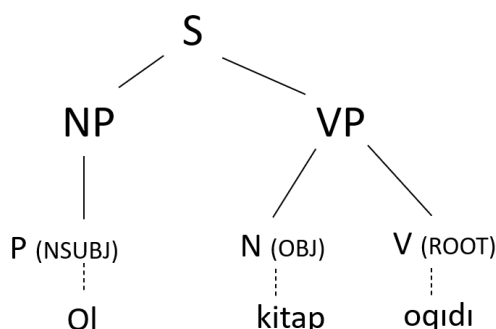
ID	Word	Lemma	POS	Head	Relation
1	Ol	Ol	PRON	3	Nsubj
2	kitap	kitap	NOUN	3	Obl
3	oqidi	oqi	VERB	0	ROOT

Bu misolda «Ol kitap oqidi» jumlasining lingvistik tahlilini ko'rishimiz mumkin. Jadvaldagi «relation» belgisi gap bo'laklarini bildirsa, «head» gap bo'laklari qaysi so'zga tobe ekanligini ko'rsatadi.

Lekin qoraqalpoq tilida sodda gaplarning bundan boshqa ham bir qancha turlari mavjud bo'lib, ularni qoraqalpoq tili grammatikasini hisobga olgan holda tahlilini ishlab chiqish maqsadga muvofiq bo'ladi. Dastlab sintaktik teglar [Elov B., 2024: 111] va tobelik belgilari (dependency labels) ro'yxati bilan tanishib chiqsak:

NP – otli birikma
 ADVP – ravishli birikma
 S – gap (darak gap)
 SINV – inversiyaga uchragan gap
 CONJP – bog'lovchi birikma
 VP – fe'lli birikma
 ROOT – kesim
 NSUBJ – ega
 OBL – to'ldiruchi
 AMOD – aniqlovchi
 ADVMOD – hol

Sintaktik tahlilning vazifasi gap tuzilishini tahlil qilish bo'lib, gap (S) eng avval egaga tobe so'zlar (NP) va kesimga tobe so'zlar (VP) guruhlariga ajratib olinadi [Abjalova M., 2020: 92]. So'ng bu guruhlar ichidagi tobe bo'laklar ham tahlil qilinadi. Qoraqalpoq tili uchun sintaktik analizatorni ishlab chiqishda Dependency Tree, Directed Acyclic Graph, Constituency Tree, Link Grammar metodlari yoki Universal Dependencies loyihasidan foydalanish mumkin. Bu vositalar orqali gap bo'laklari orasidagi bog'liqlik, bosh bo'lak va unga tobe bo'laklarning tobelik daraxti ishlab chiqiladi. Bunday tahlil turlaridan tilning grammatik xususiyatlari va qo'yilgan vazifani hisobga olgan holda qo'llash maqsadga muvofiq bo'ladi. Masalan:



Sodda gap xabarning alohida sintaktik birligi bo'lib, u bir predikativ birlikdan iborat. Qoraqalpoq tilida sodda gaplar an'anaviy o'rganish bo'yicha tuzilish xususiyatlari tarkibi, gap bo'laklarining munosabati va boshqa belgilariga ko'ra bir necha turlarga bo'linadi [Dáwletov A, Dáwletov M, Qudaybergenov M., 2009: 39]. Bosh bo'laklarning ishtirokiga ko'ra bir bosh bo'lakli va ikki bosh bo'lakli gaplar deb katta ikki guruhga bo'lib o'rganiladi.

Ikki bosh bo'lakli gap: Men universitetke bardim.

1 Men _ PRON _ Case=Nom 2 nsubj
 2 bardim _ VERB _ Tense=Past 0 ROOT
 3 universitetke _ NOUN _ Case=Dat 2 obl

Bunday natijani olish uchun NLP kutubxonalaridan foydalanish mumkin. Ammo hozirgi kunda sintaktik tahlilni amalga oshirishga yordam beruvchi kutubxonalarga (spaCy, Stanza, NLTK) qoraqalpoq tili qo'shilmagan. Shunga qaramasdan pythonda parser kodini yarim avtomatik tarzda yoki Rule-Based Approach orqali funktsiya ko'rinishida yaratish mumkin:

```
from collections import defaultdict
```

```
def karakalpak_parser(sentence):
# Grammatik ma'lumotni saqlash uchun lug'at
parsed = defaultdict(dict)
```

```
# Gapni so‘zlarga ajratish
words = sentence.split()

for word in words:
    if word == "Men":
        parsed[word] = {
            "pos": "PRON", # olmosh
            "case": "nom", # bosh kelishik
            "role": "nsubj" # ega
        }
    elif word == "universitetke":
        parsed[word] = {
            "pos": "NOUN", # ot
            "case": "dat", # jo‘nalish kelishigi (-ke)
            "role": "obl" # hol
        }
    elif word == "bardim":
        parsed[word] = {
            "pos": "VERB", # fe‘l
            "tense": "past", # o‘tgan mahal
            "person": "1", # I bet
            "role": "ROOT" # kesim
        }
```

return parsed

```
# Consolega chiqarish
sentence = "Men universitetke bardim"
analysis = karakalpak_parser(sentence)
for word, info in analysis.items():
    print(f"{word}: {info}")
```

Natija:

```
Men: {'pos': 'PRON', 'case': 'nom', 'role': 'nsubj'}
universitetke: {'pos': 'NOUN', 'case': 'dat', 'role': 'obl'}
bardim: {'pos': 'VERB', 'tense': 'past', 'person': '1', 'role': 'ROOT'}
```

Bir bosh bo‘lakli gapni tahlil qilish uchun ham yuqoridagi usuldan foydalanish mumkin:
 Universitetke bardim.

1. bardim _ VERB _ Tense=Past 0 ROOT
2. universitetke _ NOUN _ Case=Dat 1 obl

Sodda gaplar ikkinchi darajali bo‘laklarning ishtirokiga qarab yoyiq va yig‘iq gap deb ikki turga bo‘linadi. **Yoyiq gapning** tahliliga misol keltirsak: Ulgili oqiwshı qızıqlı kitap oqımaқта.

1. Ulgili — ADJ — 2 — amod1
2. oqıwshı — NOUN — 5 — nsubj
3. qızıqlı — ADJ — 4 — amod2
4. kitap — NOUN — 5 — obj
5. oqımaқта — VERB — 0 — ROOT

Bunda 3-ro‘yxatda berilgan sonlar so‘zning qaysi a‘zo bilan tobe bog‘lanishda turganligini ko‘rsatsa, amod belgisi aniqlovchilarni ifodalab kelgan.

Shuningdek, qoraqalpoq tilida qo'shmalangan (qospalangán) sodda gaplarning ham bir necha turlari bor, shulardan bir xil a'zoli (teng bo'lakli) sodda gapga misol keltirsak: Ol kitap, dápter hám ruchka satıp aldı.

1. Ol — PRON — 7 — nsubj
2. kitap — NOUN — 7 — obj
3. , — PUNKT — 2 — punkt
4. dápter — NOUN — 7 — obj
5. hám — CCONJ — 4 — cc
6. ruchka — NOUN — 7 — obj
7. satıp aldı — VERB — 0 — ROOT

Qo'shmalangan (qospalangán) gap oddiy gaplarga qaraganda murakkabroq bo'lganligi uchun ularni tahlil qilishda bir qancha qiyinchiliklar paydo bo'ladi. Xususan, tinish belgilari, yordamchi so'zlar ham alohida tokenga bo'linadi, lekin ular gap bo'lagi vazifasini bajarmaydi. Shuningdek, gap bo'laklariga bo'linmaydigan yoki gap bo'lagi vazifasini bajarib kelmaydigan so'zlarni tahlil qilishda ham ayrim to'siqlar yuzaga kelishi mumkin.

Xulosa qilib aytganda, gap turlarini tahlil qilishda sintaktik analizatorning ishlab chiqilishi ko'pgina muammolarni hal qilishga yordam beradi. Bu vazifani amalga oshirishda yechimi topilishi qiyin bo'lgan masalalarni inobatga olgan holda har tomonlama o'ylab yondashilsa maqsadga muvofiq bo'ladi.

Foydalanilgan adabiyotlar:

1. Abdurahmonova N.Z, Ótemisov A.Z. Kompyuter lingvistikasina kirisiw. Nókis – 2024.
2. Abjalova M, Tahrir va tahlil dasturlarining lingvistik modullari. Toshkent – 2020.
3. Берфорт Б, Билбро Р, Охеда Т. Прикладной анализ текстовых данных на Python. Санкт-Петербург – 2019.
4. Dáwletov A, Dáwletov M, Qudaybergenov M. Házirgi qaraqalpaq ádebiy tili. Sintaksis. Nókis – 2009.
5. Elov B, O'zbek tili korpusini sintaktik teglash masalasi. «Kompyuter lingvistikasi: muammolar, yechim, istiqbollar» (Xalqaro ilmiy-amaliy konferensiya). № 01.2024.
6. Po'latov S, Muhammedova S, Kompyuter lingvistikasi. Toshkent – 2008.
7. <https://universaldependencies.org/ru/dep/>

RASMIY (INSTITUTSIONAL) DISKURS TUSHUNCHASI VA UNING LINGVISTIK XUSUSIYATLARI

Yadgarova Elnora Abdujapbarovna

Aniq va ijtimoiy fanlar universiteti

1-bosqich Lingvistika -o'zbek tili yo'nalishi magistranti

Annotatsiya: Ushbu maqolada rasmiy (institutsional) diskurs tushunchasining lingvistik mohiyati, uning shakllanish omillari va asosiy xususiyatlari tahlil qilinadi. Diskursga oid zamonaviy yondashuvlar, turli fanlar kesimida berilgan ta'riflar, shuningdek, lingvistikadagi qo'llanishi, ayniqsa, badiiy va rasmiy kontekstlarda tadqiq etiladi. Asosan tilshunoslik sohasida olib borilgan nazariy qarashlar, G'arb va Sharq olimlarining ilmiy ishlari asosida ushbu atamaning turli ko'rinishlari yoritiladi. Shuningdek, diskursning ijtimoiy vazifasi, kommunikativ xususiyatlari va rasmiy institutsional muhitda aks etish mexanizmlari bayon qilinadi. O'zbekiston tilshunosligida olib borilgan tadqiqotlar, xususan, Sh. Safarov, A. Pardaev kabi olimlarning fikrlari tahlil markazida bo'ladi. Maqolada rasmiy diskursning og'zaki va yozma shakllari, ularning farqlanishi va kommunikativ vosita sifatidagi o'rni xolisona baholanadi. Bundan tashqari, xalqaro va milliy tadqiqotlar asosida diskursni o'rganishning dolzarb yo'nalishlari ko'rib chiqiladi.

Kalit so'zlar: rasmiy diskurs, institutsional nutq, lingvistik xususiyatlar, kommunikativ jarayon, matn va nutq, diskurs turlari, og'zaki va yozma diskurs, Sh. Safarov, Z. Harris, Teun van Deyk, tilshunoslik.

Abstract: This article analyzes the linguistic essence of the concept of official (institutional) discourse, its formation factors, and its main features. It explores modern approaches to discourse, definitions given in various fields, and its application in linguistics, particularly in literary and official contexts. The study is based on theoretical views in the field of linguistics and examines various perspectives on this term as presented by Western and Eastern scholars. The article also discusses the social functions of discourse, its communicative characteristics, and how it manifests in formal institutional settings. Research conducted in Uzbek linguistics, especially by scholars such as Sh. Safarov and A. Pardaev, is examined in detail. The article provides an objective analysis of oral and written forms of official discourse, their differences, and their roles as communicative tools. Moreover, it reviews current directions in discourse research based on both international and national studies.

Keywords: official discourse, institutional speech, linguistic features, communicative process, text and speech, types of discourse, oral and written discourse, Sh. Safarov, Z. Harris, Teun van Dijk, linguistics.

Аннотация: В данной статье анализируется лингвистическая сущность понятия официального (институционального) дискурса, факторы его формирования и основные характеристики. Рассматриваются современные подходы к дискурсу, определения, данные в различных областях, а также его применение в лингвистике, особенно в литературном и официальном контекстах. Исследование основано на теоретических взглядах в области языкознания и охватывает различные трактовки термина, представленные западными и восточными учёными. Также в статье раскрываются социальные функции дискурса, его коммуникативные особенности и механизмы реализации в институциональной среде. Отдельное внимание уделяется исследованиям, проводимым в узбекской лингвистике, в частности, работам Ш. Сафарова и А. Пардаева. Объективно анализируются устные и письменные формы официального дискурса, их различия и роль как коммуникативного средства. Кроме того, на основе международных и национальных исследований рассматриваются актуальные направления изучения дискурса.

Ключевые слова: официальный дискурс, институциональная речь, лингвистические особенности, коммуникативный процесс, текст и речь, типы дискурса, устный и письменный дискурс, Ш. Сафаров, З. Харрис, Тюн ван Дейк, лингвистика.

Rasmiy (institutsional) diskurs bugungi lingvistik tadqiqotlar doirasida dolzarb mavzulardan biri hisoblanadi. Diskurs tushunchasining o'zi ilk bor G'arbiy tilshunoslikda, xususan Z. Xarris tomonidan 1952 yilda ilmiy muomalaga kiritilgan bo'lsa-da, uning mazmuni va qo'llanilish doirasi keyingi yillarda sezilarli darajada kengaygan. Bu atama dastlab faqat til birliklarining matndagi bog'liqligini ifodalash uchun ishlatilgan bo'lsa, hozirgi kunda diskurs til va jamiyat, til va hokimiyat, til va ijtimoiy institutlar o'rtasidagi murakkab aloqalarni ifodalovchi konsept sifatida qaralmoqda. Ayniqsa, rasmiy yoki institutsional diskurs atamasi ma'lum ijtimoiy kontekstda, odatda davlat, qonunchilik, ta'lim, sog'liqni saqlash va boshqa rasmiy muassasalar faoliyatida yuzaga keladigan nutq shakllarini ifodalaydi.

Rasmiy diskurs, eng avvalo, aniq kommunikativ vazifaga xizmat qiluvchi, barqaror struktura va ifoda shakliga ega bo'lgan nutq shakli sifatida e'tirof etiladi. Unda til vositalarining rasmiy, neytral, aniq va me'yorga solingan ishlatilishi ustuvor ahamiyatga ega. Bunday diskurs jamiyatdagi siyosiy, huquqiy va ma'muriy institutlarning faoliyatida, ular tomonidan qabul qilinadigan qarorlar, bayonotlar, qonunlar, protokollar va boshqa rasmiy hujjatlarda kuzatiladi. Diskurs, albatta, matn va nutqdan iborat bo'lib, u faqat lingvistik birliklardan emas, balki ularning ortida turgan kontekst, maqsad, kommunikantlar o'rtasidagi munosabatlar va ijtimoiy muhit kabi ekstralingvistik omillarni ham o'z ichiga oladi.

Zamonaviy lingvistik yondashuvlarga ko'ra, diskursni faqat til birliklari yig'indisi deb emas, balki ularni voqelantiruvchi va tushuntiruvchi ijtimoiy hodisa sifatida ham qarash lozim. Bu jihatdan qaraganda, rasmiy diskurs davlat institutlari faoliyatida tildan qanday foydalanilishini, til orqali hokimiyat qanday ifodalanishini va qanday vositalar yordamida ma'lumotlar yetkazilishini tahlil qilish imkonini beradi. Jumladan, rasmiy nutqning sintaktik tuzilishi, leksik boyligi, funksional-stilistik xususiyatlari va kommunikativ niyatlari ushbu diskurs tahlilida muhim o'rin tutadi.

Diskursga doir nazariyalar orasida G'arb tilshunosligida Teun van Deyk, Norman Fairclough, Deborah Schiffrin kabi olimlarning ishlari alohida o'rin egallaydi. Ularning ta'kidlashicha, diskurs doimiy ravishda ijtimoiy tuzilmalar bilan o'zaro aloqada bo'lib, ular orqali ijtimoiy hokimiyatning ifodalanish shakllarini tahlil qilish mumkin bo'ladi. O'zbekiston tilshunosligida ham diskurs tadqiqotlari faol olib borilmoqda. Sh.Safarov, A.Pardaev kabi olimlar diskurs va matn muammolarini, ularning tafovutli va umumiy tomonlarini keng tahlil qilganlar. Aynan shu tadqiqotlar orqali diskursning faqat og'zaki yoki yozma shakl emasligi, balki inson tafakkuri va ijtimoiy ong mahsuli sifatida baholanishi lozimligi ilgari surilmoqda.

Shuningdek, diskurs turlarining ko'pligi, ularning muayyan kontekstga bog'liqligi va ijtimoiy vazifasi bu tushunchaning keng qamrovli ekanini ko'rsatadi. Jumladan, badiiy, siyosiy, baholovchi, ilmiy va rasmiy diskurslar mavjud bo'lib, ularning har biri o'ziga xos strukturaviy va funksional belgilariga ega. Ushbu maqolada aynan rasmiy diskursning lingvistik xususiyatlari, ularni tashkil qiluvchi birliklar, kommunikativ uslublar, pragmatik maqsadlar va kontekstual shart-sharoitlar asosida tahlil qilinadi.

Maqolaning maqsadi — rasmiy (institutsional) diskursni lingvistik yondashuv asosida chuqur o'rganish, unga oid mavjud qarashlarni tahlil qilish, shuningdek, rasmiy muloqotdagi tildan foydalanish mexanizmlarini aniqlab berishdan iborat.

Diskurs tushunchasi hozirgi lingvistikada keng tadqiq etilayotgan va turli fanlararo yondashuvlarni o'z ichiga oluvchi murakkab hodisadir. Ayniqsa, rasmiy (institutsional) diskurs ijtimoiy munosabatlarning lingvistik ko'rinishi sifatida e'tirof etiladi. U til orqali ijtimoiy institutlar faoliyatining huquqiy, siyosiy va ma'muriy jihatlarini ifodalash, me'yorlashtirish va boshqarishga xizmat qiladi. Z.Xarris tomonidan 1952 yilda ilgari surilgan "diskurs tahlili" tushunchasi rasmiy nutq birliklarining mazmuniy va kontekstual o'ziga xosliklarini tahlil qilish

imkonini yaratgan. Teun van Deyk, Fairclough kabi tilshunoslar esa diskursni hokimiyat va ijtimoiy kuchlar bilan bevosita bog'laydi, ya'ni rasmiy nutq orqali ijtimoiy nazorat va boshqaruv amalga oshirilishini ko'rsatadi.

Rasmiy diskurs, o'z tabiatiga ko'ra, doimo institutlashtirilgan kommunikatsiya shaklidir. U aniq formulalar, huquqiy yoki tashkiliy normativlarga asoslangan iboralar, yuqori darajadagi rasmiylik va neytrallikni saqlaydigan leksik-semantik vositalar orqali ifodalanadi. Masalan, qonun hujjatlari, farmonlar, rasmiy bayonotlar, protokollar, xatlar va boshqa ma'muriy matnlar rasmiy diskursning asosiy namunalari andir. Ular nafaqat axborot yetkazish, balki ijtimoiy tartibni shakllantirish, ongli ravishda muomala chegaralarini belgilash funksiyalarini bajaradi.

Rasmiy diskursga xos bo'lgan lingvistik birliklar, ayniqsa, sintaktik tuzilmalar, murakkab iboralar, passiv konstruksiyalar va abstrakt leksik birliklar orqali ajralib turadi. Masalan, "amaldagi qonunchilikka muvofiq ravishda", "yuqori tashkilot ko'rsatmasi asosida", "hududiy vakolatli organ" kabi iboralar kommunikativ neytrallikni ta'minlaydi, shu bilan birga tilni me'yorlashtirishga xizmat qiladi. Bu xususiyatlar rasmiy diskursni boshqa nutq turlaridan ajratib turadi va uni ijtimoiy institutlar bilan bevosita bog'laydi.

Ushbu diskurs turining asosiy belgilari qatoriga ijtimoiy mavqega asoslangan ierarxik aloqa, standartlashtirilgan til vositalari, noma'lumlik darajasi yuqori bo'lgan mavhum konstruksiyalar kiradi. Bu esa rasmiy nutqni o'ziga xos uslubiy qatlamga aylantiradi. Diskurs bu yerda nafaqat tilning qo'llanilishi, balki ma'lumotning qanday shaklda, kim tomonidan, kim uchun va qanday maqsadda yetkazilishiga ham bevosita bog'liq bo'ladi. Ayni holatda "institutsional diskurs" atamasi kontekstual va kommunikativ cheklovlar bilan shakllanadi, bu esa tilning ijtimoiy funksiyasini kengaytiradi.

O'zbekiston tilshunosligida ham bu mavzuga oid izlanishlar olib borilmoqda. Jumladan, Sh.Safarov o'z tadqiqotlarida matn va diskurs muammolarini qiyosiy o'rganadi. Unga ko'ra, og'zaki va yozma nutq o'rtasidagi tafovut, informativ va kommunikativ mazmun uyg'unligi, diskursning real muloqotdagi semantik xossalari chuqur tahlil etilishi kerak. Bundan tashqari, A.Pardaev fikricha, diskurs – bu insoniy faoliyat, unda lisoniy va nolisoniy vositalar o'zaro uyg'unlikda voqelashadi. Bu esa rasmiy diskursni faqat matn shaklida emas, balki kommunikativ jarayon sifatida o'rganish zarurligini anglatadi.

Diskursning ijtimoiy funksiyasi ham alohida ahamiyat kasb etadi. Rasmiy diskurs davlat va jamiyat o'rtasidagi muloqot maydoni sifatida xizmat qiladi. U orqalik hokimiyat vakillari o'z siyosiy pozitsiyalarini bildiradi, aholiga mo'ljallangan axborot tarqatiladi, ijtimoiy tartib saqlanadi. Bu esa rasmiy nutq vositasida amalga oshirilayotgan lingvistik boshqaruvni, ya'ni tildan hokimiyat quroli sifatida foydalanishni anglatadi. Misol sifatida rasmiy telekanallardagi nutq, hukumat farmoyishlari, muassasa ichki hujjatlari, notarial bayonotlar va boshqa shu kabi matnlar tahlil qilinadi. Ularning barchasida muloqot subyektlari o'rtasida asimmetrik aloqa shakllanadi, ya'ni hokimiyat pozitsiyasi ustunlik qiladi.

Tilshunos olim Muxamadiyeva Nodira o'z maqolasida rasmiy diskursni ekstralingvistik omillar bilan bog'liq tarzda ko'rib chiqadi. Unga ko'ra, diskurs nafaqat til birliklarining yig'indisi, balki ma'lum ijtimoiy, psixologik, madaniy va kognitiv omillarni o'z ichiga oluvchi hodisadir. Demak, har qanday diskursni, xususan, rasmiy nutqni baholashda faqat lingvistik yondashuv emas, balki kommunikatsion va ijtimoiy-lingvistik asoslar ham inobatga olinishi zarur.

Rasmiy diskurs til madaniyati va ijtimoiy identifikatsiya vositasi sifatida ham faol ishlatiladi. Rasmiy nutqda ishlatiladigan iboralar, ohang va ifoda shakllari muayyan ijtimoiy guruh, tashkilot yoki davlat tizimining o'ziga xosligini, madaniy yondashuvini va muloqot etikasi darajasini namoyon etadi. Shu bois rasmiy diskurs o'zida milliy, siyosiy va madaniy qadriyatlarni mujassam etgan lingvistik konstruksiya hisoblanadi.

Rasmiy (institutsional) diskurs tilshunoslikning muhim va murakkab sohalaridan biri bo'lib, u nafaqat lingvistik birliklarni o'rganish, balki ularning kommunikativ, ijtimoiy va madaniy kontekstdagi ifodalanish shakllarini tahlil qilishga xizmat qiladi. Diskurs zamonaviy lingvistika doirasida ijtimoiy hodisa sifatida qaralmoqda, bunda til faqat aloqa vositasi emas, balki ijtimoiy

ong, hokimiyat va madaniy qadriyatlarni shakllantiruvchi asosiy omillardan biri hisoblanadi. Ayniqsa, rasmiy diskursda til me'yorlashtirilgan, strukturalashgan va maqsadga yo'naltirilgan tarzda qo'llaniladi, bu esa uni boshqa nutq turlaridan ajratib turadi.

Rasmiy diskurs davlat institutlari, huquqiy tizim, ta'lim, sog'liqni saqlash, axborot siyosati kabi sohalarida asosiy kommunikatsiya vositasi bo'lib xizmat qiladi. U orqali ijtimoiy boshqaruv, axborot tarqatish va ma'lumot yetkazish amalga oshiriladi. Undagi tildan foydalanish qoidalari, leksik va sintaktik birliklarning tanlanishi ijtimoiy muvozanat, madaniy tartib va rasmiy etikaga bo'ysunadi. Bu holat rasmiy diskursni ijtimoiy funktsiyaga ega bo'lgan til hodisasiga aylantiradi. Shuningdek, u me'yoriy iboralar, rasmiy formulalar, passiv shakllar va mavhum tushunchalarning ustunligi bilan ajralib turadi, bu esa uning kommunikativ neytralligini ta'minlaydi.

Diskursning zamonaviy tadqiqotlarida uning ekstralingvistik jihatlariga — ya'ni sotsiologik, psixologik, madaniy va kognitiv omillariga alohida e'tibor qaratilmoqda. Rasmiy diskursda bu omillar rasmiylik darajasini belgilaydi, muloqot taraflari o'rtasidagi ierarxik munosabatlarni ifodalaydi va til orqali hokimiyatning ifodalanishiga xizmat qiladi. Rasmiy diskursda muloqot ko'pincha bir tomonlama va asimmetrik xarakterga ega bo'lib, axborot yetkazuvchi subyektning ustun pozitsiyasini aks ettiradi.

Tilshunos Sh.Safarov, A.Pardaev va boshqa olimlarning fikrlariga ko'ra, diskurs nafaqat matnli yoki nutqiy birlik sifatida, balki insoniy faoliyatning ko'rinishi sifatida ham talqin qilinishi kerak. Shuningdek, Muxamadiyeva Nodira va Saidova Muxayyo kabi zamonaviy tadqiqotchilar rasmiy diskursning lingvistik qirralarini chuqur o'rganib, uning zamonaviy jamiyatdagi rolini aniqlashtirishga xizmat qilmoqdalar. Bunday tadqiqotlar shuni ko'rsatadiki, rasmiy diskurs tilshunoslikda doimiy e'tibor markazida bo'lishi lozim bo'lgan murakkab ijtimoiy-lingvistik hodisadir.

Xulosa qilib aytganda, rasmiy (institutsional) diskurs nafaqat tilning kommunikativ imkoniyatlarini aks ettiradi, balki jamiyatdagi ijtimoiy tuzilmalarni mustahkamlovchi va ma'lumotlar oqimini tartibga soluvchi lingvistik vosita sifatida ham alohida ilmiy va amaliy ahamiyatga ega. Uning chuqur va tizimli o'rganilishi tilshunoslikning turli yo'nalishlarida yangi yondashuvlar, tadqiqot metodologiyalari va tahlil uslublarini shakllantirishga xizmat qiladi.

Foydalanilgan adabiyotlar

1. Ablakulova O. U. (2021). Diskurs tushunchasi va uning tadqiqi. Academic Research in Educational Sciences, 2(11), 1167–1172. DOI: 10.24412/2181-1385-2021-11-1167-1172.
2. Muxamadiyeva N. A. (2023). Tilshunoslikda diskurs tushunchasi va uning tahlili. Badiiy diskurs. Raqamli texnologiyalar davrida tillarni intensiv o'qitishning psixologik-pedagogik jihatlar. Respublika ilmiy-amaliy anjumani materiallari. <https://doi.org/10.5281/zenodo.7992890>
3. Saidova M. U., Nadirxanovna N. A. (2024). Diskurs tushunchasi tahlili va tadqiqi. Miasto Przyszłości, 44, 136–138.
4. Fairclough N. (2001). Language and Power (2nd ed.). London: Longman.
5. van Dijk, T. A. (1997). Discourse Studies. SAGE Publications.

SINTAKTIK ANNOTATSIYALANGAN KORPUSNING AHAMIYATI

Rashidova Muxlisa Ilxomovna

O'zMU Kompyuter lingvistikasi yo'nalishi 1- bosqich magistrant

E-mail: tolibboyevamuxlisa@gmail.com

Annotatsiya: ushbu maqolada sintaktik annotatsiyalangan korpuslarning tilshunoslik va kompyuter lingvistikasi sohalaridagi o'rni, zamonaviy sun'iy intellekt tizimlaridagi ahamiyati tahlil qilinadi. O'zbek tilida sintaktik annotatsiyalangan korpus yaratishning dolzarbligi haqida so'z boradi.

Tayanch so'zlar: sintaktik annotatsiya, kompyuter lingvistikasi, sintaktik analiz, korpus, korpus annotatsiyasi.

Annotation: This article analyzes the role of syntactically annotated corpora in the fields of linguistics and computational linguistics, and their importance in modern artificial intelligence systems. The paper addresses the relevance of developing a syntactically annotated corpus for the Uzbek language.

Keywords: syntactic annotation, computational linguistics, syntactic analysis, corpus, corpus annotation.

Аннотация: В данной статье анализируется роль синтаксически аннотированных корпусов в лингвистике и компьютерной лингвистике, а также их значение в современных системах искусственного интеллекта. Обсуждается актуальность создания синтаксически аннотированного корпуса для узбекского языка.

Ключевые слова: синтаксическая аннотация, компьютерная лингвистика, синтаксический анализ, корпус, аннотация корпуса.

Bugungi kunda axborot texnologiyalari rivojlanib borgan sari unga bo'lgan ehtiyoj ham oshib bormoqda. Kompyuter texnologiyalari shiddat bilan rivojlanib borayotgan vaqtda har qanday fan bilan o'zaro hamkorlik yo'lga qo'yilmoqda. Bunda kompyuter texnologiyalari til rivojiga o'z ta'sirini ko'rsatib natijada korpus lingvistikasi, kompyuter leksikografiyasi, mashina tarjimai, tahrirllovchi dastur, nutqiy sintezator, matnlarni tasniflash kabi bir qancha yo'nalishlar vujudga keldi.

So'nggi yillarda tabiiy tilni qayta ishlash (NLP – Natural Language Processing) sohasida kompyuter lingvistikasi hamda korpus lingvistikasi asosiy tayanch vositalardan biriga aylandi.

Til ijtimoiy hodisa ekan undagi muammolarni yechishda axborot kommunikatsion texnologiyalarning o'rni beqiyos. Bu borada Kompyuter lingvistikasi fani muhim ahamiyat kasb etadi. Lingvistik bilimlarni rivojlantirish bo'yicha dunyo tajribasiga tayanish hamda erishilgan yuksak natijalardan oqilona foydalanish talab etiladi [Abduraxmonova, 2021; 6].

Kompyuter lingvistikasi – bu til va kompyuter texnologiyalarining kesishgan nutqtasida joylashgan fan bo'lib, tabiiy tillarni kompyuter orqali qayta ishlash, tahlil qilish va tushunishga xizmat qiladi. Kompyuter lingvistikasi uchun korpus juda muhim hisoblanadi. Korpus bu sohaning asosiy poydevori desak ham bo'ladi. Korpus lingvistikasining nazariy ildizlari Noam Chomskiy ilgari surgan generativ tilshunoslik bilan bog'liq. Matnlarni to'plam shaklida avtomatik tahlil qilish korpus lingvistikasi taraqqiyoti uchun muhim voqelik bo'ldi [Abduraxmonova, 2023; 12].

Korpusda annotatsiya – bu matndagi elementlarga qo'shimcha axborot yoki belgi qo'shish jarayoni hisoblanadi. Annotatsiyalash muhim jarayon hisoblanadi.

Matnni annotatsiyalash quyidagi jihatlariga ko'ra tasniflanadi:

1. Ektralingvistik (metarazmetka) – matn (muallifi, nashr yili, nashriyot nomi, yili, janri, mavzusi) va uning muallifi haqida ma'lumot.

2. Matn tuzilishi – sarlavha, satrboshi, gap, so‘z shakli.

3. Lingvistik sathiga ko‘ra:

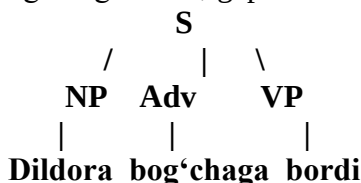
1. Morfologik (POS-tagging) razmetka.
2. Sintaktik razmetka – bunda so‘zlar o‘rtasidagi sintaktik aloqalar va gap turlariga ko‘ra ma’lumotlar keltiriladi.
3. Semantik razmetka – matnda ifodalangan tushunchalarning narsa-predmet va narsa-predmet bo‘lmagan nomlari, faoliyat turlari hamda semantik munosabatlari tavsiflanadi;
4. Anaforik razmetka – matnning biror elementi boshqa matnda anglashilgan mazmun bilan aloqadorlikda aniqlanadi va lingvistik jihatdan tahlil qilinadi;
5. Prosodik razmetka – urg‘u, ritm, mantiqiy urg‘u va boshqlariga ko‘ra matn lingvistik jihatdan ifodalanadi[Боярский, 2013;28].

Sintaktik annotatsiyalangan korpus tilshunoslikda va tabiiy tilni qayta ishlash (NLP) sohasida juda muhim resurs hisoblanadi. Sintaktik annotatsiyalar til strukturasi, jumla tuzilishi, so‘zlararo munosabatlarni chuqur o‘rganish imkonini beradi. Sintaktik annotatsiya – matnlar ustida sintaktik darajadagi ma’lumotlarni, ya’ni so‘zlar o‘rtasidagi grammatik bog‘lanishlarni ko‘rsatish jarayoni bo‘lib, bu korpuslar mashinali tarjima, avtomatik xulosa chiqarish, so‘z birikmalari tahlili, nutqni tahlil qilish kabi ko‘plab amaliy jarayonlarga asos bo‘ladi [Nivre,2020;4035].

So‘z birikmalarini xorijda XVIII asrning boshlarida tilshunoslar tomonidan o‘rganish boshlangan. XX asrga kelib esa sintaksis bo‘yicha juda ko‘p ishlar qilingan edi.

Tabiiy tilni qayta ishlash uchun tillarning lingvistik xususiyatlaridan muhimi bu sintaksisdir. Sintaktik strukturalarni o‘rganish tillarning xususiyatidan kelib chiqib, turli nazariyalarni yaratilishiga sabab bo‘lgan. Tobelanish nazariyasi uzoq tarixga ega bo‘lib, Panini grammatikasi va arab tilshunoslarining ilmiy qarashlarida o‘z ifodasini topgan [Abduraxmonova, 2021; 55]. Dunyodagi mashhur sintaktik annotatsiyalangan korpuslar orasida Penn Treebank (Ingliz tili) korpus hisoblanadi. Bu korpuslar millionlab so‘z birliklarini o‘z ichiga oladi va morfologik, sintaktik, ba’zida semantik belgilanishlar bilan boyitilgan. Korpus – bu til materiallarining tizimli to‘plami bo‘lib, u elektron shaklda saqlanadi va tilni o‘rganish yoki tahlil qilish uchun ishlatiladi. Sintaktik annotatsiyalashda esa gap hamda so‘z birikmalari ustida tahlil qilish amalga oshiriladi. Sintaktik annotatsiyalash gapning tuzilishini, ya’ni so‘zlarning o‘zaro sintaktik bog‘liqliklarini aniqlash jarayonini ifodalaydi. Sintaktik analiz (parsing) esa matndagi muayyan sintaktik birlik va strukturalarni avtomatik tarzda analiz qiladi[Abduraxmonova, 2021; 154].

Bu bosqich esa NLP uchun juda muhim hisoblanadi. Sintaktik tahlil (parsing) uchun turli metodlar ishlab chiqilgan .Ulardan biri Noam Chomskiy yaratgan daraxtsimon tahlil qilish metodidir. Bu gap yoki jumla tuzilishini daraxt ko‘rinishida tasvirlash metodidir. Sintaktik tuzilishni izohlash daraxt shaklini eslatganligi sabali ingliz tilida u Treebank deyiladi. Bu usulda jumladagi har bir so‘z va ibora o‘z grammatik funksiyasiga ko‘ra daraxtning tugunlariga joylashtiriladi. Masalan, Dildora bog‘chaga bordi, gapida



S – gapni, NP – otli frazani, ADV – hol uchun berilgan belgini, VP – fe‘lli frazani ifodalaydi. Korpus yordamida so‘z birikmasi hamda gap turlarini tahlil qilish insonga ancha qulayliklar tug‘diradi .

Bugungi kunda o‘zbek tili uchun yirik, ommaviy va to‘liq sintaktik annotatsiyalangan korpus hali mavjud emas. Sintaktik annotatsiyalangan korpusning mavjud bo‘lishi tilning sintaktik tuzilishini chuqur o‘rganishni, NLP texnologiyalarini yaratishni, qisqa qilib aytganda, milliy NLP texnologiyalarining rivoji uchun juda muhim qadam bo‘lishini aytishimiz mumkin.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Kompyuter lingvistikasi (darslik), Toshkent: Nodirabegim, 2021. - B. 391
2. Nivre, J. et al Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. *LREC*. . (2020).Abduraxmonova N. Korpus lingvistikasi(darslik),Toshkent:GlobeEdit.2023Боярский К. К. Введение в компьютерную лингвистику. – Санкт-Петербург, 2013. – С. 28.
3. Uzbek Treebank Project: <https://github.com/uzbeknlp/uzbek-treebank>
4. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, 11(19), 9117.
5. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).

DUNYO TILSHUNOSLIGIDA SOTSIOLEKTLARNING TASNIFLANISHI VA O'RGANILISHI

Tursunova Nigora Komil qizi

Annotatsiya: Zamonaviy tilshunoslikda til va jamiyat o'zaro chambarchas bog'liq bo'lib, bu bog'liqlik sotsiolingvistika fanida chuqur o'rganiladi. Aynan shu sohada **sotsiolektlar** – ya'ni ijtimoiy guruhlar orasida shakllanadigan til variantlari muhim o'rin tutadi. Sotsiolektlar kishilarning kasbi, yoshi, jinsi, ijtimoiy mavqei, mintaqaviy joylashuvi yoki etnik mansubligiga qarab farqlanadi. Ushbu maqolada dunyo tilshunosligidagi mashhur sotsiolektlar, ularni o'rganish uslublari va til taraqqiyotiga ta'siri yoritiladi.

Kalit so'zlar: sotsiolingvistika, sotsiolektlar, empirik tadqiqotlar, til strukturasi, milliy til, dialekt, kasbiy leksika.

Abstract: In modern linguistics, language and society are intricately interconnected, a relationship deeply explored in sociolinguistics. Within this field, sociolects—language varieties emerging among social groups—hold significant importance. Sociolects vary according to factors such as profession, age, gender, social status, regional location, or ethnic affiliation. This article examines prominent sociolects in global linguistics, their research methodologies, and their impact on language development.

Keywords: sociolinguistics, sociolects, empirical research, language structure, national language, dialect, professional lexicon.

Аннотация: В современной лингвистике язык и общество тесно взаимосвязаны, и эта связь глубоко изучается в социолингвистике. В рамках этой области социолекты — языковые варианты, возникающие среди социальных групп — занимают важное место. Социолекты различаются в зависимости от таких факторов, как профессия, возраст, пол, социальный статус, региональное расположение или этническая принадлежность. В данной статье рассматриваются известные социолекты в мировой лингвистике, методы их исследования и влияние на развитие языка.

Ключевые слова: социолингвистика, социолекты, эмпирические исследования, структура языка, национальный язык, диалект, профессиональный лексикон.

Til insoniyatning eng muhim ijtimoiy hodisalaridan biri bo'lib, u orqali nafaqat muloqot amalga oshadi, balki ijtimoiy guruhlar, sinflar, madaniyat va shaxsiy identitet ham ifodalanadi. Aynan shu nuqtai nazardan, zamonaviy sotsiolingvistika fanida sotsiolektlar muhim obyekt sifatida o'rganilmoqda. Sotsiolektlar — bu ma'lum bir ijtimoiy guruh tomonidan foydalaniladigan til shakli bo'lib, u kishining kasbi, jinsi, yoshi, madaniyati yoki ijtimoiy mavqeiga qarab farqlanadi [Bourdieu P., 1991: 20].

Sotsiolekt termini lingvistikada XX asrning ikkinchi yarmida paydo bo'lgan. U sotsio – jamiyatga bo'lgan munosabat va dialekt so'zlaridan tashkil topgan bo'lib, rus tilidagi “социальный диалект” (ijtimoiy dialekt) birikmasining qisqargan shaklidir. Sotsiolekt – umumiy ijtimoiy xususiyatlariga ko'ra birlashgan ayrim guruhlarining lisoniy ehtiyojlarini qondirish uchun nutq amaliyotida ishlab chiqilgan til shakli. Standart tildan farqli o'laroq, sotsiolektda ba'zi o'rinlarda chetga chiqish mumkin. Til ham xuddi tashqi ko'rinish singari sinflar haqida guvohlik beradi. Sotsiolekt verbal xulq va ijtimoiy ahvolning muvofiqligini ifodalaydi [Elova D., Safarova G., 2021: 64].

Dastlab bu tushuncha William Labov tomonidan keng ommaga tanitilgan bo'lib, u o'zining New York City aksentlari haqidagi tadqiqotlarida til va ijtimoiy sinf o'rtasidagi bog'liqlikni aniqlagan [Labov W., 1972: 215]. Tilshunos Basil Bernstein esa sotsiolektlarni “elaborated code”

va “restricted code” tarzida farqlab, ijtimoiy sinflar tilining struktura va ekspressivligidagi tafovutlarni tahlil qilgan [Bernstein B., 1971: 156].

Dunyo tilshunosligida sotsiolektlarni o‘rganish va tadqiq qilishda bir qator olimlar o‘z hissalarini qo‘shganligi bilan ajralib turadi. Quyida ularning ayrimlari haqida qisqacha ma’lumot berib o‘tishni joiz deb topdik.

1. William Labov – Amerikalik tilshunos olim bo‘lib, u sotsiolektlarni o‘rganishga katta hissa qo‘shgan. U o‘zining “sotsial tilshunoslik” nomli tadqiqotlari bilan tanilgan bo‘lib, tilning ijtimoiy kontekstda qanday ishlashini tahlil qilgan.

2. Peter Trudgill – Britaniyalik tilshunos, u ham sotsiolektlarni o‘rganishga qaratilgan ko‘plab tadqiqotlar olib borgan va ijtimoiy omillarni tilga ta’sirini tahlil qilgan.

3. Basil Bernstein – Ijtimoiy kommunikatsiya va ta’lim sohalarida ishlagan olim, uning “ishtirok etuvchi” va “rasmiy” kommunikatsiya tili (sotsiolektlar) haqida qarashlari ham bugungi dunyo tilshunosligi uchun nihoyatda muhimdir.

4. Eckert Penelope – Sotsiolingvistika sohasida faoliyat yurituvchi olimlardan biri bo‘lib, u yosh guruhlar va ularning til uslublari orasidagi farqlarni o‘rganadi.

Bu olimlar sotsiolektlarni tadqiq qilishda asosiy nazariyalar va metodologiyalarni ishlab chiqqanlar. Shuningdek, turli mamlakatlarda va madaniyatlarda ham sotsiolektlarni o‘rganishga qaratilgan ko‘plab tadqiqotlar mavjudki, ular ushbu mamlakatlarning ijtimoiy ahvoli, madaniyati va rivojlanish darajasiga ko‘ra turlichadir.

Shunga ko‘ra sotsiolektlarni o‘rganishda turli metodlardan foydalaniladi:

- **Empirik tadqiqotlar** — suhbat, so‘rovnoma, kuzatuv.
- **Fonetik-fonologik tahlil** — talaffuzdagi farqlar.
- **Korpus lingvistikasi** — katta hajmdagi yozma va og‘zaki matnlar asosida statistik tahlil.

Diskursiv va sotsiometrik metodlar — til va ijtimoiy mavqe o‘rtasidagi bog‘liqlikni tahlil qiladi [Labov W., 1972: 231].

Sotsiolektlar orqali kishilar jamiyatdagi o‘rnini, qaysidir guruhga mansubligini ifodalaydi. Ular musiqada (rep, tango), adabiyotda, kino va internet tilida ham keng qo‘llaniladi [Bourdieu P., 1991: 200]. Ayniqsa yoshlar orasidagi leksik o‘zgarishlar tez rivojlanadi. Biroq ba’zida sotsiolektlar stigmatizatsiyaga uchraydi, ya’ni “past tabaqaning tili” deb noto‘g‘ri talqin qilinishi mumkin.

“Sotsiolekt” termini turli-tuman, bir biriga o‘xshamaydigan, ayni paytda umumiy xususiyatlarga ega bo‘lgan lisoniy shakllarni ifodalash uchun qo‘llanilishi to‘g‘risida yuqorida ham aytib o‘tdik. Mazkur shakllar ijtimoiy chegaralangan kishilar guruhining kommunikativ ehtiyojlariga xizmat qiladi. Sotsiolektlar kommunikatsiyaning butun bir tizimini ifodalamaydi. Chunonchi, u so‘zlar, so‘z birikmalari, sintaktik konstruksiyalar, urg‘u xususiyatlari va h.k. ko‘rinishidagi nutqning o‘ziga xosligidir [Elova D., Safarova G., 2021: 65].

Sotsiolektlar asosini milliy tildan deyarli farq qilmaydigan lug‘at va grammatika tashkil qiladi. Masalan, jinoyatchilar argosida o‘ziga xos ma’no, xususan, metaforik ma’no kasb etgan so‘zlar mavjud: ment – “militsoner” etap qilish – “qamoqqa jo‘natish”, narkota – “narkotik, giyohvand moddalar bilan shug‘ullanish” va h.k. Biroq mazkur so‘zlarning turlanishi, tuslanishi va gapdagi bog‘lanishi umumtil qoliplari va qoidalariga muvofiq amalga oshiriladi. Hatto, so‘zlar ham umumtil leksikasiga mansub bo‘ladi. Quyida ingliz tilida mavjud ba’zi sotsiolektlarning turlari bilan qisman tanishib o‘tamiz:

1. African American Vernacular English (AAVE)

AAVE — afroamerikaliklar orasida keng tarqalgan ingliz tilining o‘ziga xos ijtimoiy varianti bo‘lib, u fonetik, grammatik va stilistik xususiyatlari bilan ajralib turadi. Misol uchun:

- “He be working” — bu doimiy harakatni bildiradi.
- “I ain’t got no money” — ikki karra inkor ishlatilgan [Rickford J. R., 1999: 89].

William Labov AAVEni grammatik tizimi bor, o‘z ichki qoidalariga ega tildan biri sifatida ko‘rsatadi [Labov W., 1972: 254].

2. Cockney English

Cockney — Londonning ishchi sinfiga mansub boʻlgan sotsiolekt boʻlib, oʻziga xos “rhyming slang” bilan mashhur: “dog and bone” = phone, “apples and pears” = stairs [Hickey, 2004: 476]. Fonetik xususiyatlaridan biri bu “th” tovushining “f” yoki “v” bilan almashinishidir (“think” → “fink”).

3. Estuary English va Received Pronunciation (RP)

Estuary English janubiy Angliya hududlarida keng tarqalgan va RP bilan Cockney oʻrtasidagi oʻrta holat deb qaraladi. RP esa yuqori tabaqa inglizlarining talaffuzi boʻlib, koʻpincha “BBC English” deb ham ataladi [Kerswill P., 2003: 15].

4. Teen Slang (Oʻsmirlar leksikasi)

Internet, pop madaniyat va ijtimoiy tarmoqlar orqali shakllangan soʻzlar: “lit” (zoʻr), “ghosting” (aloqani birdan uzish), “salty” (xafa) [Crystal D., 2001: 344]. Bular yoshlar oʻrtasidagi identitetni koʻrsatadigan lingvistik vositalardir.

Sotsiolektlar tilshunoslik uchun juda muhim ahamiyatga ega. Ular tilning ijtimoiy kontekstda qanday ishlashini, jamiyat ichidagi turli guruhlarining tilni qanday oʻzgartirishi va rivojlantirishini oʻrganishga yordam beradi.

Quyidagi jihatlar sotsiolektlarning tilshunoslikdagi ahamiyatini koʻrsatadi:

1. Ijtimoiy identifikatsiya: Sotsiolektlar orqali odamlar oʻzlarini ijtimoiy guruhlariga mansubligini ifodalaydilar. Bu esa ularning kimligini, madaniyatini va tarixini aks ettiradi.

2. Tilning oʻzgaruvchanligi: Sotsiolektlar tildagi oʻzgarishlarni koʻrsatadi. Ularning tahlili orqali til qanday qilib vaqt oʻtishi bilan moslashayotganini va yangi soʻzlar, frazeologik birliklar yoki grammatik tuzilmalar qachon paydo boʻlishi mumkinligini tushunish mumkin.

3. Madaniyat va tarixi: Sotsiolektlar maʼlum bir madaniy yoki tarixiy kontekstdan kelib chiqishi mumkin, bu esa tildagi maʼnolarni va ifodalash usullarini yanada chuqurroq anglashga yordam beradi.

4. Til va kuch dinamikasi: Sotsiolektlar orqali kuch dinamikalarini (masalan, kuchli yoki zaif ijtimoiy guruhlar orasidagi munosabatlarni) koʻrish mumkin. Qanday qilib baʼzi tillar yoki dialektlar boshqalardan ustun kelishini tushunish uchun sotsiolektlarni tahlil qilish zarur.

5. Tillararo muloqot: Sotsiolektlarni oʻrganish orqali turli tillar va madaniyatlar oʻrtasidagi muloqotni yaxshilash mumkin, bu esa globallashtirish jarayonida muhim ahamiyatga ega.

Umuman olganda, sotsiolektlarning tahlili tilshunoslikda juda muhim boʻlib, u tilning ijtimoiy hayotdagi rolini anglashga yordam beradi.

Xulosa qilib shuni taʼkidlash joizki, sotsiolektlar tilning ijtimoiy funksiyalarini chuqur tahlil qilishda muhim ahamiyat kasb etadi. Ular orqali tilshunoslar nafaqat til strukturasini, balki ijtimoiy qatlamlar, tengsizliklar va madaniy tafovutlarni ham aniqlay oladi. Shuningdek, sotsiolektlar til taraqqiyotining asosiy harakatlantiruvchi kuchlaridan biri boʻlib, jamiyat tilidagi xilma-xillikni boyitadi va shu sababli ham dunyo tilshunosligida hanuzgacha oʻrganilishda davom etmoqda.

Adabiyotlar roʻyxati:

1. Bernstein B. *Class, Codes and Control: Theoretical Studies towards a Sociology of Language*. London: Routledge. 1971. – P. 268
2. Bourdieu P. *Language and Symbolic Power*. Cambridge, MA: Harvard University Press. 1991. – P. 302
3. Crystal D. *Language and the Internet*. Cambridge: Cambridge University Press. 2001. – P. 427
4. Elova D., Safarova G. *SOTSIOLINGVISTIKA (Oʻquv-uslubiy qoʻllanma)*, Toshkent – 2021. – B 100
5. Hickey R. *Legacies of Colonial English: Studies in Transported Dialects*. Cambridge: CUP. 2004. – P. 712
6. Kerswill P. *Dialect levelling and geographical diffusion in British English*. In *Britain and Cheshire, Sociolinguistics*: 2003. –12-17-p.

7. Labov W. Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press. 1972. – P. 344
8. Rickford J. R. African American Vernacular English: Features, Evolution, Educational Implications. Malden, MA: Blackwell. 1999. – P. 42

BADIIY MATNLARDA ARGO VA JARGONLARNING FUNKSIONAL XUSUSIYATLARI

Topildiyeva Feruza

O'zMU Kompyuter lingvistikasi
va amaliy tilshunoslik kafedrasida katta o'qituvchisi

Rasulberdiyeva Zulfizar Doniyor qizi

O'zMU Jurnalistika va o'zbek filologiyasi fakulteti 4-kurs talabasi

Annotatsiya: Mazkur maqolada badiiy asarlarda argo va jargon leksikasining funksional xususiyatlari tahlil qilinadi. Argo va jargonlar badiiy matnlarda qo'llanilishi asarda obraz yaratish, sotsiokulturologik kontekstni ifodalash va asardagi hazil, kinoya, tabu kabi mavzularni yumshoq ifodalash singari turli kommunikativ stilistik vazifalarni bajaradi. Maqolada bu birliklarning adabiy tildagi o'rni, ularning asarda jonli obraz yaratishdagi ahamiyati nazariy va amaliy misollar asosida ochib berilgan.

Kalit so'zlar: Argo, jargon, badiiy matn, stilistik vosita, ijtimoiy til, harbiylar jargon, tibbiyot jargon, o'g'rilar argosi, jinoyatchilar argosi, personaj nutqi.

Аннотация: В данной статье анализируются функциональные особенности арго и жаргонной лексики в художественных произведениях. Использование арго и жаргона в художественных текстах выполняет различные коммуникативно-стилистические функции, такие как создание образов, выражение социокультурного контекста, смягчение тем, связанных с юмором, иронией, табу и т.п. В статье теоретически и на практических примерах рассматривается место этих единиц в литературном языке и их значение в создании живых образов в произведении.

Ключевые слова: Сленг, жаргон, литературный текст, стилистический прием, социальный язык, военный жаргон, медицинский жаргон, воровской жаргон, уголовный жаргон, речь персонажей.

Annotation: This article analyzes the functional characteristics of argot and jargon vocabulary in literary works. The use of argot and jargon in literary texts serves various communicative and stylistic functions, such as creating vivid characters, expressing sociocultural context, and softening themes like humor, irony, and taboos. The article examines the role of these units in the literary language and their importance in creating lively images through both theoretical and practical examples.

Keywords: Slang, jargon, literary text, stylistic device, social language, military jargon, medical jargon, thieves' slang, criminals' slang, character speech.

Adabiyot – bu hayotning oynasi. Unda insoning his tuyg'ulari, ijtimoiy munosabatlari va madaniy qadriyatlari til vositasida ifodalanadi. Badiiy matnlar tilning boyligini va ifodalilik imkoniyatlarini aks ettiruvchi muhim vosita hisoblanadi. Badiiy asar muallifi asar qahramonlarini jonli va ishonarli qilish uchun turli til qatlamlariga, jumladan argo va jargonlarga murojaat qilishi mumkin. Argo va jargonlardan foydalanish orqali personajlarning ijtimoiy, psixologik va madaniy xususiyatlarini yorqinroq bayon qiladilar. Argo va jargonlar nutqning mazmunini boyitish, real hayotga yaqinlashtirish, obrazlarni jonlantirish va ma'yan ijtimoiy guruhlarining til xususiyatlarini aks ettirishda muhim rol o'ynaydi. Argo va jargonlarning badiiy asarlardagi funksional xususiyatlariga batafsil bayon qilishdan oldin argo va jargon atamalar nima? Qanday holatlarda ishlatiladi? Kimlar tomonida qo'llaniladi? kabi savollarga to'xtalsak. Argo va jargon atamasi bir –biriga yaqin atama hisoblanadi. Shunday bo'lsa-da, bu ikki terminning bir-biridan ajratib turuvchi bir qancha xususiyatlari mavjud.

Argo atamasi fransuz tilidagi “argot” soʻzidan olingan boʻlib, maʼlum sotsiol guruhning oʻziga xos tili. Argo nutqda kommunikatsiya predmetini sir tutish maqsadida qoʻllaniladi.¹²⁶ Argolar qiziqishlari, mashgʻulotlari, yoshlari bir xil boʻlgan maʼlum tor doiradagi guruh vakillari tomonida ishlatiladigan soʻzlardir. Masalan, oʻgʻrilar nutqida: urkagan – tajribali oʻgʻri, mandarin – tilla; mahbuslar nutqida: krokodil – qaychi, lebed – geroin kabi soʻzlar misol boʻla oladi.

Jargon maʼlum tabaqa vakillari yokida maʼyyan bir soha va guruhlariga xos, tashqaridan tushunish qiyin boʻlgan maxsus atamalarni anglatadi. Masalan harbiylar jargoni, talabalar jargoni, yoshlar jargon, tibbiyot jargoni, kompyuter jargonlari boʻlishi mumkin. Talabalar jargoniga chizmoq – qochmoq, quloq – chaqimchi, yopmoq – sessiyani tugatmoq kiradi. Harbiy jargoniga esa diskodeka – oshxona naryadi, ded – xizmatdan qaytishi yaqinlashgan asakarga nisbatan shu kabi soʻzlar qoʻllaniladi. Jargonlar tarkibi boshqa tildan oʻzlashgan soʻzlar bilan ham boyib borishi mumkin.

Badiiy matnda argo va jargonlarning oʻrni, ularning funksiyalari va estetik qiymati nuqtai nazaridan juda muhimdir. Badiiy asarlarda bu birliklar obraz yaratish, individuallashtirish, realizm va tabiiylikni yaratish, ekspressivlik, emotsionallikni kuchaytirish singari vazifalarni yuzaga chiqarishda yordam beradi. Bundan farqli ravishda, jargon koʻproq professional yoki texnik atmosferalarda ishlatiladi, bu esa muayyan kasb yoki faoliyat guruhlariga oid maxsus atamalar toʻplamini anglatadi. Badiiy asar ichida jargonning kiritilishi, asarning mavzu va mohiyatini yanada chuqurlashtirishga xizmat qiladi va oʻquvchiga ushbu professional muhitni his qilish imkoniyatini beradi. Masalan, shifokorlar, muhandislar yoki sanʼatkorlar bilan bogʻliq terminologiyalar, qahramonning shu soha bilan bogʻliqligini ochib beradi. Badiiy matnlarda argo qoʻllanishi - bu tilning ham obrazli, ham badiiy kuchini oshiruvchi muhim xususiyatlardan biridir. Argolar, maʼlum bir guruhga yoki qatlamga xos til ifodalari sifatida, maʼlum bir kontekstlarda individual va maʼnaviy boʻshliqlarni toʻldirishi, shuningdek, oʻquvchini qiziqtirishi orqali ijodiy ifodaning intensivligini taʼminlashda katta rol oʻynaydi. Badiiy adabiyotda argo soʻzlar koʻpincha qahramonlarning xarakterini ochish, muhitni taʼriflash va dramatik taʼsirni kuchaytirish maqsadida keng qoʻllaniladi. Ushbu elementlar oʻquvchini hikoyaga jalb qilish va turli figurativ maʼno qatlamlarini ochish orqali qiziqarli, bevosita va chiroyli bir badiiy obraz yaratishda ahamiyatlidir. Masalan, argo tilida ifodalangan soʻzlar, oʻquvchiga muayyan qahramonning ichki olamini anglashga yordam beradi va ijodkorning niyatini yanada kuchliroq his etishga imkon yaratadi. Argo koʻpincha taʼsirchan va qiziqarli obrazlar yaratadi, shuningdek, maʼlum bir davr yoki muhit haqida foydali maʼlumotlarni kitobxonga yetkazadi.

Badiiy asarda qoʻllanilgan argo va jargonlarni tahlil qilib chiqsak:

- Oqidanmi, qizilidanmi, qaysisidan olay deymen, ayang oʻrgilsin?
- Keragi yoʻq, aya. Bilasz-ku, ichmayman.¹²⁷

Bu gapdagi “oqidanmi”, “qizilidanmi” soʻzlari argo hisoblanadi. Yaʼni ichuvchilar tilida aroqqa nisbatan oq soʻzi, vinoga nisbatan qizil soʻzlari qoʻllaniladi. Bu soʻzlar adabiy tilda mavjud va ichuvchilar argonida bu soʻzlarga boshqa maʼno yuklanyapti.

“Shu qizingni menga oshira qolsang-chi? Barbir, sen uni... eplay olmaysan”¹²⁸. Badiiy asardagi yuqoridagi jumlada gapiruvchi bu yerda boshqa odamni mensimaslik, uni oʻzini qiziqtirgan insonni ushlab qoʻlishga noliyiq deb hisoblash orqali oʻzini ustun qoʻyisha urunmoqda. Gapdagi “oshira qolsang-chi” jumlasiga argodir. Aslida “oshirmoq” feʼli biror narsani yuqoriga koʻtarish, berib yuborish maʼnosini bildiradi. Bu feʼlning kontekstdagi maʼnosi esa “oshira qolsangchi” – “qizingni menga bersangchi” degan argo va erkin ohangdagi taklif. Bu soʻzlar koʻcha bolalari, yigitlari, yoshlar orasida norasmiy suhbatda qoʻllaniladi. Shuningdek, bu jumla asarda ifodalanishi bilan qahramonning ijtimoiy maqomi haqida yaʼni uning oddiy koʻcha yigiti ekani va obrazning hazilmutoyiba, tanbehli, loqayd yoki tajavuskor feʼl-atvori ochildi.

¹²⁶ R.Sayfullayeva, Hozirgi oʻzbek adabiy tili, Toshkent., 2010

¹²⁷ O.Yoqubov, Billur qandillar, Toshkent., 2024. 9 bet

¹²⁸ O.Yoqubov, Billur qandillar, Toshkent., 2024. 23bet

Muayyan ijtimoiy guruhga taalluqli jargonlarni yozuvchilar obraz yaratish, muayyan muhitni ifodalash, aholining u yoki bu guruhining nutqini namoyon etish, urf-odatlarini ko'rsatish maqsadida qo'llaydilar. Shuningdek, Abdulla Qodiriyning hajviy asarlari "Toshpo'lat tajang nima deydi" va "Kalvak mahzumning xotira daftarida" hajviylarida jargonlardan mohirona foydalanilgan. Shular qatori G'afur G'ulomning "Shum bola" qissasida ijtimoiy argolar, Tog'ay Murod asarlarida kurashga xos jargonlar, Tohir Malikning "Shaytanat", "Alvido, bolalik" kabi asarlarida kriminal muhit tasvirlangani uchun o'g'rilar, jinoiy guruhlarining argo va jargonlarini uchratishimiz mumkin. "Shum bola" asaridagi argo va jargonlarning funksional xususiyatlarini ko'rib chiqsak:

"Mana bundan tomteshar chiqadi, kissavur bo'lolmaydi."¹²⁹ Ko'rinib turibdiki, bu gap o'g'rilar tilidan aytilgan. Bu gapdagi "tomteshar" so'zi o'g'rilar argonida ishlatilib, uyga tushib o'g'rilik qiladigan o'g'rilar nazarda tutiladi. Shu kabi so'zlar qo'llanilishi obrazning o'g'ri ekanligi va uning bu boradan ancha yaxshi tajribasi borligini ochiqlab kelgan.

"Ertada-kech ko'cha changitib hammaning joniga tegib, kampirlarda qarg'ish eshitib, o'spirinlardan kaltak yeb, sandiroqlab yuradigan uvinto'da bekorchi bolalarmiz."¹³⁰ Bu gapdan anglash mumkinki, nut egasi o'ziga tanqidiy qaraydigan bola, shahar ko'chasida hayot kechirayotgan yosh bola ekanligi tushiniladi. "uvinto'da" atamasi ko'cha bolalari ishlatadigan argo hisoblanadi. Kontekstda betarib ma'nosida qo'llanilgan.

Artur Konan Doylning "Sherlok Holms haqida hikoyalari" asarida ham turli argo va jargonlarni uchratamiz. "O'ladigan kuni, tushdan so'ng Ronald Bagatet klubida bir robberdan vist o'ynaganligi aniqlangan."¹³¹ Bu gapdagi "vist" klassik kart o'yini va bu so'z kart o'yinlar olamida ishlatiladi. Mazkur gapdagi "robber" so'zini izohlaydigan bo'lsak, bu so'z ingliz tilida o'g'ri ma'nosini bildiradi. Lekin asarda o'g'ri ma'nosida kelmagan. Ya'ni bu yerda kart o'yinining bir bosqich ma'nosida kelgan. "Vist", "robber" so'zlari qimorbozlar jargonida qo'llaniladigan so'zlardir. "Moriartining shaykasi ustidan bo'lgan suddan so'ng unga a'zo bo'lgan ikkita xavfli bandit erkinlikda qolgan." Ushbu gapda "shayka" so'zi jinoyatchilar argosida ishlatiladigan so'z. "Shayka" to'da, guruh ma'nosini anglatadi. Bu so'zlarni asarda bayon qilinishi qimorboz va jinoyatchilarni yanada jonli va ta'sirli gavdalandirish uchun yordam beradi.

Badiiy matnlarda argo va jargonlarning qo'llanilishi o'zbek adabiyotida tilning yanada tabiiylshuvi, personajlar psixologik holatini chuqur ifodalash, ijtimoiy haqiqatni ko'rsatish yo'lida muhim vosita sifatida xizmat qiladi. Ular asarga hayotiylik, ekspressivlik, realistic tus bag'ishlaydi. Shu bois, argo va jargonlari faqatgina "qo'pol" va "ruxsat etilmagan" til elementi sifatida emas, balki badiiy-estetik qiymatga ega stilistik vosita sifatida baholash kerak.

Foydalanilgan adabiyotlar:

1. A.Konan Doyl, Sherlok Holms haqida hikoyalari, "O'qituvchi" NMIU., 2018
2. H.Jomolxonov, Hozirgi o'zbek adabiy tili, Toshkent – Talqin – 2005
3. O.Yoqubov, Billur qandillar, Toshkent., 2024. 9 bet
4. R.Sayfullayeva, Hozirgi o'zbek adabiy tili, Toshkent., 2010
5. Z.Taxirov, O'zbek tilining amaliy stilistikasi, Toshkent., 2020
6. Z. Xolmonova, Tilshunoslikka kirish, Toshkent., 2007
7. G'.G'ulom, Shum bola, Toshkent., 2018
8. Sh. Rahmatullayev, Hozirgi o'zbek adabiy tili,

¹²⁹ G'.G'ulom, Shum bola, Toshkent., 2018

¹³⁰ G'.G'ulom, Shum bola, Toshkent., 2018

¹³¹ A.Konan Doyl, Sherlok Holms haqida hikoyalari, "O'qituvchi" NMIU., 2018

TERMINOLOGIK TEZAVURUS YARATISHDAGI NAZARIY VA AMALIY YONDASHUVLAR BIOLOGIYA SOHASI MISOLIDA

Mukhammadaliev Farangizkhon

Biologiya fakulteti 4-kurs talabasi
O'zbekiston Milliy universiteti

Annotatsiya: Ushbu maqolada terminologik tezavrus tushunchasining nazariy asoslari va biologiya fanida uni yaratishning amaliy jihatlarini tahlil qilingan. Tezavrus – bu terminlar o'rtasidagi semantik aloqalarni ifodalovchi tizim bo'lib, ilmiy va o'quv jarayonida axborotning mantiqiy bog'liqlik asosida tuzilgan bazasini yaratishda muhim vosita hisoblanadi. Biologik terminlarning sinonimiya, antonimiya va iyerarxik munosabatlarini asoslash orqali maqolada zamonaviy lingvistik yondashuvlar yoritilgan.

Kalit so'zlar: tezavrus, biologik termin, semantik aloqa, sinonimiya, antonimiya, iyerarxiya.

Annotation: This article analyzes the theoretical foundations of the concept of a terminological thesaurus and the practical aspects of its creation in biology. A thesaurus is a system that expresses semantic relationships between terms and is an important tool for creating a logically connected database of information in the scientific and educational process. The article highlights modern linguistic approaches by substantiating the synonymy, antonymy, and hierarchical relationships of biological terms.

Keywords: thesaurus, biological term, semantic relation, synonymy, antonymy, hierarchy.

Аннотация: В статье анализируются теоретические основы концепции терминологического тезауруса и практические аспекты его создания в биологической науке. Тезаурус — это система, отражающая семантические связи между терминами, которая является важным инструментом для создания логически связанной базы данных информации в научном и образовательном процессе. В статье освещаются современные лингвистические подходы, обосновывающие синонимию, антонимию и иерархические отношения биологических терминов.

Ключевые слова: тезаурус, биологический термин, семантическая связь, синонимия, антонимия, иерархия.

Hech kimga sir emaski, yaqin ikki asrda biologiya sohasi doirasida bir qancha yangi yo'nalishlar, yangi fanlar paydo bo'ldi. Ilm-fan bilan shug'ullanishni istagan, biologiya sohasiga qiziqqan o'quvchilar, talabalar, o'qituvchilar, tadqiqotchilarning ushbu sohaga oid terminlarning tezavrusiga ehtiyojlari borligi ma'lum.

Prezidentimiz Shavkat Mirziyoyev ta'kidlaganlaridek: Bugun O'zbekistonimiz "Milliy tiklanishdan – milliy yuksalish sari" degan bosh tamoyil asosida taraqqiyotning yangi, yanada yuksak bosqichiga ko'tarilmoqda. Mamlakatimizda amalga oshirilayotgan keng ko'lamlı islohotlar jarayonida davlat tilining hayotimizdagi o'rni va nufuzi tobora oshib bormoqda [O'zbekiston Respublikasi Prezidentining "O'zbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan oshirish chora-tadbirlari to'g'risida"gi, 21.10.2019 yildagi PF-5850-son Farmoni]. Shu jihatdan, ilmiy asoslangan yangi so'z va atamalarni iste'molga kiritish, zamonaviy atamalarning o'zbekona muqobillarini yaratish, tilimizning nufuzini oshirishda asosiy tamoyillardan biri hisoblanadi.

Tezavrus yaratish biologiya sohasini kompyuter lingvistikasi tarmog'i bilan bog'lab, ilmiy materiallarga tezkor kirishni ta'minlaydi. O'zbek tilining maqomini oshirishga qaratilgan chora-tadbirlar yuzasidan, ushbu lug'atning yaratilishi, tilimizni keng ko'lamda o'rganishga, tilimiz imkoniyatlaridan to'la foydalanishga yordam beradi.

Tezaurus tushunchasining mohiyati. “Tezaurus” (thēsauros – qadimgi yunon tilida “xazina”, “boylik”) tushunchasi ilk bor Aristotel va Platon davrida ishlatilgan bo‘lsa-da, lingvistikada bu tushuncha XX asr o‘rtalaridan boshlab faol qo‘llanila boshlandi. Tezauruslar, avvalo, ma‘lumotlarni strukturaviy tartibga solish, sinonimik tarmoqlarni aniqlash, fanlararo terminlarning bog‘liqligini izohlash kabi bir qator vazifalarni bajaradi.

R. Temmerman (2000) sotsiokognitiv yondashuv doirasida tezauruslarni til birliklarining kontekstual, prototipik va ko‘p ma‘noli tabiatini aks ettiruvchi vosita sifatida talqin qiladi. U tezaurusni qat‘iy strukturaviy rasmiy model emas, balki o‘zgaruvchan semantik maydon sifatida tushunishni taklif etadi [Rita T., 2000. – P 131-135.].

Tezaurusning lingvistik tizimdagi roli. Til tizimi, xususan ilmiy til – bu faqat so‘zlar majmuasi emas, balki ular orasidagi munosabatlar tizimidir. Bu tizimda so‘zlarning semantik darajasi, ularning boshqa birliklar bilan munosabatlari, giperonimik va giponimik qatlamlari muhim o‘rin tutadi.

F. Kurbanova (2023) tezaurus lug‘atlar haqida shunday deydi: “Har bir so‘zning sinonimi, antonimi, etimologiyasi, omonimi va u bilan birika oladigan so‘zlar qatori ochib beriladi... Bu lug‘atning elektron formatdagisi esa insonlarning vaqtini tejash, bir so‘z haqida to‘lig‘icha ma‘lumot olish imkonini beradi” [Kurbanova F. S. 2023. – B 59-60.].

Tezauruslar orqali:

- terminlar fanlararo bog‘liqliklar asosida guruhlanadi;
- har bir termin uchun ma‘no doirasi, talaffuz shakli, grammatik xususiyatlari beriladi;
- sinonim, antonim va semantik “yaqinlik” bog‘lamlari aniqlanadi.

Bu yondashuv ayniqsa o‘quvchilarning fanlararo tafakkurini shakllantirishda, ilmiy matnlarni tahlil qilishda va axborotlarni tizimlashtirishda beqiyos ahamiyatga ega.

Tezaurusning axborot tizimidagi o‘rni. Axborot jamiyatida insonlar har kuni ulkan hajmdagi ma‘lumotlarga duch keladi. Ular orasidan keraklisini aniqlash, sintez qilish va tuzilmaga solish – faqat semantik indeksatsiya tizimlari yordamida amalga oshiriladi. Bu yerda kompyuter lingvistikasi va sun‘iy intellekt asosida ishlovchi tezauruslar markaziy o‘ringa chiqadi.

Ph. Royle va boshqalar (2009) ta‘kidlaganidek, tezauruslar ko‘p hollarda korpus tahlili asosida tuziladi va bu tahlillar matndagi terminlar orasidagi ko‘rinmas semantik aloqalarni aniqlashga xizmat qiladi. Ayniqsa, fanlararo foydalaniladigan terminlar uchun tezaurus ma‘no farqlarini ajratish vositasi bo‘lib xizmat qiladi [Phaedra R. and collaborators. 2009. – P 523-524.].

Shuningdek, axborot tizimlarida:

- matnlarni avtomatik kategoriyalash;
- ilmiy maqolalarning konseptual tahlili;
- o‘quvchi bilimni baholovchi intellektual sistemalar yaratishda tezauruslar asos bo‘lib xizmat qiladi.

Terminologik tezauruslar fan tilining asosiy birliklari – terminlarni semantik, morfologik va konseptual tizimga solishning eng mukammal shakllaridan biridir. Ular terminlar o‘rtasidagi munosabatlarni vizual, tartibli va iyerarxik ko‘rinishda ifodalash orqali fan tili va tuzilmasini chuqurroq tushunishga xizmat qiladi. Shuning uchun tezaurus yaratishda yondashuvlar nazariy asoslanishi bilan birga, amaliy ahamiyatga ham ega hisoblanadi.

Nazariy yondashuvlar: tizimlilik va semantik strukturaviylik.

Tezaurus yaratishda nazariy asos sifatida klassik terminologiya, kognitiv lingvistika, kompyuter lingvistikasi va sotsiokognitiv yondashuvlar birgalikda qo‘llaniladi.

1. An‘anaviy (klassik) yondashuv. Bu yondashuvga ko‘ra, har bir termin bir ma‘noli, qat‘iy aniqlangan, iyerarxik joylashgan bo‘lishi kerak. Bunda:

- terminlar giperonim–giponim qatlamlarda joylashtiriladi;
- har bir atama intensional va extensional ta‘rif bilan beriladi;
- tezaurus ontologik sxemalarga asoslanadi.

2. Sotsiokognitiv yondashuv (Temmerman). R. Temmerman (2000) tezaurusni tirik, kontekstual va dinamik struktura deb hisoblaydi. Unga ko‘ra:

- terminlar ijtimoiy-kognitiv muhitda shakllanadi;
- prototiplar asosida tasniflanadi;
- sinonimiya, polisemiyaning mavjudligi inkor etilmaydi, balki kognitiv o'zgaruvchanlik sifatida e'tirof etiladi.

Bu yondashuv biologik terminologiya uchun ayniqsa mos keladi, chunki biologiya fanida ko'plab atamalar kontekstga qarab o'z semantikasini o'zgartiradi [Rita T. 2000. – P 132-135.].

Amaliy yondashuvlar: strukturaviy va texnologik asoslar

Tezaurusni amalda tuzishda quyidagi bosqichlar muhim hisoblanadi:

1. Terminlar korpusini yig'ish. Biologik darsliklar (5–11-sinf), ilmiy maqolalar, lug'atlar asosida terminlar tanlab olinadi.

2. Terminlar strukturaviy tahlili:

- semantik tahlil: asosiy ma'no, sinonimlar, antonimlar, giponimlar, giperonimlar;
- morfologik tahlil: so'z shakli, urg'usi, affikslar;
- konseptual tahlil: qaysi tushunchani bildiradi, qanday jarayonga oid.

3. Iyerarxik tartibga solish. Terminlar daraxtsimon ko'rinishda, yuqoridan pastga qarab mantiqiy guruhlarda joylashtiriladi. Bu:

- o'quvchilar uchun tushunarli strukturani yaratadi;
- ilmiy tarjimada aniqlik beradi;
- axborot tizimlarida avtomatik ajratishni osonlashtiradi.

4. Vizual modellash. Tezaurus diagrammalari, tarmoqlar, sxemalar asosida tuziladi.

Bunda semantic network (ma'no tarmog'i) konsepsiyasidan foydalaniladi.

Fanlararo yondashuvning zarurligi. Biologik terminlar nafaqat biologiya fanida, balki **kimyo, ekologiya, informatika** fanlari bilan ham kesishgan bo'lishi sababli, tezaurusda:

- kontekst belgisi (BIO/KIM/EKO);
- variant ishlanmalar;
- fanlararo sinonimlar va ekvivalentlar berilishi lozim.

Bunday tizim o'qituvchilar, tarjimonlar, dasturchilar, talaba va tadqiqotchilar uchun universal platforma bo'lib xizmat qiladi.

Biologik terminologiyaning barqarorligi hamisha dolzarb masala bo'lib kelgan. Ba'zi atamalar turlicha shaklda yoki variantlarda qo'llaniladi, bu esa ayniqsa o'quvchilarning fan mazmunini anglashiga to'sqinlik qilishi mumkin. B. E. Antia va R. A. Kamai (2016) shunday holatlarni Nigeriyada o'rganib, "variant terminologiya" tushunchasini ilgari suradilar. Ular biologiya darsliklarida bir xil tushunchaning turlicha atamalar bilan ifodalanishi natijasida o'quvchilarning bilim olish jarayonida terminologik to'siqlarga duch kelishini ko'rsatganlar [Bassey E. A. and Richard A. K. 2016. – P 204-205.].

Biologik terminlarni tizimlashtirish, ularni tezaurus shaklida taqdim etish uchun dastlab mavjud terminologik bazalar va lug'atlar tahlil qilinadi. O'zbekiston ta'lim tizimida foydalanilayotgan lug'atlar va darsliklar terminlarning shakllanishi, izohlilik, sinonimik qatlamlarining mavjudligi jihatidan turlicha darajada tuzilgan.

1-jadval. Darsliklar va lug'atlar asosida olib borilgan tahlil natijalari.

Sinflar / Manba	Terminlar soni	Izoh berilganmi	Guruhlash mavjudmi	Sinonim/Antonim aloqalar	Lotincha shakllar
Eski 5-sinf botanika	50+	Qisman	Yo'q	Yo'q	Yo'q
7-sinf zoologiya	100+	Qisman	Qisman	Yo'q	Qisman
10-sinf biologiya	150+	Ha	Qisman	Yo'q	Ha
Yangi 5-sinf biologiya	40+	Ha	Qisman	Yo'q	Yo'q
6-sinf tabiiy fanlar	30+	Qisman	Yo'q	Yo'q	Yo'q

8-sinf odam salomatligi	80+	Ha	Ha	Yo‘q	Qisman
9-sinf biologiya	120+	Ha	Qisman	Qisman	Ha
11-sinf biologiya	200+	Ha	Ha	Yo‘q	Ha
Izohli lug‘at (2016)	1000+	Ha	Yo‘q	Yo‘q	Qisman

Xulosa: Terminologik tezaurus yaratish – bu oddiy lug‘at tuzishdan farqli ravishda mantiqiy, semantik, kognitiv va vizual strukturani birlashtirgan murakkab jarayondir. Nazariy yondashuvlar (klassik, sotsiokognitiv, korpus lingvistikasi) tezaurusga mustahkam asos bo‘lsa, amaliy yondashuvlar esa uni foydalanuvchi uchun qulay shaklda taqdim etadi. Ayniqsa biologiya fanidagi terminlar bo‘yicha tuziladigan tezauruslar o‘quv jarayonini soddalashtirish, ilmiy tafakkurni shakllantirish va axborot texnologiyalari bilan integratsiyani ta‘minlashda muhim vosita hisoblanadi.

Foydalanilgan adabiyotlar:

1. O‘zbekiston Respublikasi Prezidentining Farmoni. “O‘zbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan oshirish chora-tadbirlari to‘g‘risida” PF-5850. 21.10.2019. <https://lex.uz/docs/4543266>
2. Kurbanova F. S. (2023). Tezaurus lug‘atlar haqida. // *Tilshunoslik va amaliyot*, №2. – B. 59–60.
3. Temmerman R. (2000). *Towards New Ways of Terminology Description: The Sociocognitive Approach*. – Amsterdam: John Benjamins Publishing Company. – P. 131–135.
4. Royle P., et al. (2009). Thesaurus-based linguistic indexing in digital environments. // *Language Resources and Evaluation*, Vol. 43. – P. 523–524.
5. Antia B. E., Kamai R. A. (2016). Variant Terminology in Biology Textbooks: A Study in Nigerian Classrooms. // *International Journal of Science Education*, Vol. 38. – P. 204–205.
6. Biologiyadan qisqacha izohli lug‘at / Tahr. hay‘ati: Q. Jo‘rayev, T. Rasulov va boshq. – Toshkent: O‘qituvchi, 2016.
7. Botanika: 5-sinf uchun darslik / A. Karimov, G. Sharipova, M. Rajabov. – Toshkent: O‘qituvchi, 2009.
8. Zoologiya: 7-sinf uchun darslik / M. Xolmurodov, A. Abdurahmonov, D. Yusupov. – Toshkent: O‘qituvchi, 2012.
9. Biologiya: 10-sinf uchun darslik / I. Xolmatova, Sh. Raxmatullayev, S. Mamatqulova. – Toshkent: O‘qituvchi, 2018.
10. Odam va uning salomatligi: 8-sinf uchun darslik / Z. Xudoyberdiyeva, A. Hasanov, M. Jumanazarova. – Toshkent: O‘qituvchi, 2017.
11. Biologiya: 9-sinf uchun darslik / M. Rajabov, G. Matkarimova, I. Boymatov. – Toshkent: O‘qituvchi, 2020.
12. Biologiya: 11-sinf uchun darslik / I. Xolmatova, D. Turg‘unov, M. Po‘latov. – Toshkent: O‘qituvchi, 2021.

RAVISH SO‘Z TURKUMINI MA’NOVIY TUZILISHIGA KO‘RA GRAMMATIK POS TEGGLASH (PARTS-OF-SPEECH TAGGING) MASALASI

Karimova Zilola

ToshDO‘TAU Kompyuter lingvistikasi 1-kurs magistranti

E-mail: karimovazilola791@gmail.com

Annotatsiya. Ushbu tezisda o‘zbek tilida ravish so‘z turkumining ma’noviy tuzilishi va ularni grammatik jihatdan belgilash (POS tagging) masalalari ko‘rib chiqiladi. Matnni avtomatik qayta ishlashda ravishlarning roli, ularning sintaktik va semantik xususiyatlari hamda kompyuter lingvistikasida qo‘llanilishi tahlil qilinadi. Tadqiqotda ravishlarning turli ma’noviy turlari misolida POS teglash jarayonining afzalliklari va qiyinchiliklari yoritiladi.

Kalit so‘zlar: ravish, ma’noviy tuzilish, POS teglash, morfologik tuzilishi, tabiiy tilni qayta ishlash (NLP), ma’noviy turlari.

Abstract. This thesis examines the semantic structure of adverbs in the Uzbek language and addresses their grammatical labeling (POS tagging) issues. The study analyzes the role of adverbs in automatic text processing, focusing on their syntactic and semantic features, as well as their application in computational linguistics. By exploring various semantic types of adverbs, the advantages and challenges of POS tagging are highlighted.

Keywords: adverb, semantic structure, POS tagging, NLP, morphological structure, semantic types

Kirish. O‘zbek tilining boy grammatik tuzilishi va so‘z turkumlari tadqiqotchilar diqqat markazida bo‘lib kelmoqda. Shulardan biri – ravish so‘z turkumi, harakatning vaqt, joy, usul yoki darajasini ifodalash kabi ma’noviy vazifalarni bajaradi [1]. Zamonaviy kompyuter lingvistikasi (Computational Linguistics) sohasida har bir so‘z turkumining aniq belgilanishi, ya’ni POS teglash (Parts-of-Speech Tagging), matnni avtomatik qayta ishlashda muhim ahamiyat kasb etadi [2].

Ravishlar semantik boylikka ega bo‘lib, matnda vaqt, joy, usul, sabab va daraja kabi turli ma’no nuqtalarini ko‘rsatadi. Bu esa ularning POS teglash jarayonida alohida e’tibor talab qilinishiga sabab bo‘ladi [3]. Zero, turli NLP vazifalarida (mashinada tarjima, semantik tahlil, nutqni tanish) ravishlarning to‘g‘ri belgilanishi umumiy aniqlik darajasini oshiradi [4].

Muhokama va natijalar

Ravishlar o‘zbek tili morfologiyasida mustaqil so‘z turkumi sifatida quyidagi asosiy ma’no guruhlariga bo‘linadi :

- Payt ravishlari: kecha, bugun, ertaga, hozir;
- O‘rin ravishlari: yuqorida, pastda, oldinda, ichkarida;
- Holat ravishlari: tez, sekin, ehtiyotkorlik bilan;
- Sabab ravishlari: noiloj, ilojsiz, chorasizlikdan;
- Maqsad ravishlari: ataylab, atay, jo‘rttaga;
- Miqdor-daraja ravishlari: juda, nihoyatda, ancha, uncha.[5]

Ushbu turkumlar matn semantikasini boyitishda katta rol o‘ynaydi. Ayniqsa, NLP jarayonlarida har bir ravishning o‘ziga xos leksik-semantik xususiyatlarini aniqlash zarur.

Tabiiy tilni qayta ishlashda so‘zlar ustida avtomatik teglash jarayonini to‘g‘ri amalga oshirish juda muhim [6]. O‘zbek tilida ravishlarni POS teglash quyidagi bosqichlarni o‘z ichiga oladi:

Ravishlar bo‘yicha POS teglash bosqichlari

Tr	Bosqich	Qisqacha tavsif
1	Leksik tahlil (Tokenization)	So‘zning lug‘aviy shaklini aniqlash
2	Morfologik tahlil (Morphological analysis)	Ravishning asosiy grammatik shakllarini topish

3	Ma'noviy turini aniqlash	Payt, o'rin, holat, sabab, maqsad, miqdor-daraja
4	Teglash (Tagging)	Universal dependencies (ADV, RR)

Xalqaro miqyosda ishlab chiqilgan Universal Dependencies (UD) tizimida ravishlar odatda “ADV” teg bilan belgilangan bo‘lib, ushbu turkumning kontekstual semantikasi ham hisobga olinadi [7]. Bunda ravishlarning gapdagi sintaktik roli (masalan, fe’lni to‘ldirishi, butun gapga taalluqliligi va hokazo) ham inobatga olinadi.

Ravishlarning to‘g‘ri teglanishi mashinada tarjima, matnni semantik tahlil qilish, nutqni sintezlash, ma’lumotlarni indekslash kabi ko‘plab jarayonlarda katta ahamiyatga ega. Quyidagi jadvalda ravishlar ishtirok etadigan asosiy NLP jarayonlari keltirilgan:

NLP jarayonlarida ravishlarning ahamiyati

T/r	NLP jarayoni	Ravishlarning roli
1	Mashinada tarjima (Machine Translation)	O‘rin va payt ravishlari kontekstni aniqlashda muhim
2	Semantik tahlil (Semantic Analysis)	Sabab, maqsad va miqdor-daraja ravishlarini to‘g‘ri aniqlash, gap ma’nosini to‘g‘ri tushinish
3	Nutqni sintezlash	Holat ravishlarining intonatsion ohang va tezligini belgilash
4	Matnni indekslash (Information Retrieval)	Ravishlarni ma’no toifalariga ko‘ra to‘g‘ri indekslash

Bu jarayonlarda ravishlarni avtomatik tahlil qilish orqali matnning to‘liq semantik xaritasi yaratilishi, so‘z birikmalarining ma’nosi chuqurroq tushunilishi va tarjima jarayonlari aniqligi oshadi.

Xulosa. Shunday qilib, o‘zbek tilida ravish so‘z turkumini ma’noviy tuzilishiga ko‘ra grammatik POS teglash masalasi zamonaviy kompyuter lingvistikasi uchun dolzarb yo‘nalishlardan biri hisoblanadi. Ravishlarning turli semantik guruhlariga (vaqt, joy, usul, sabab, daraja) matnni avtomatik qayta ishlash jarayonlarida katta ahamiyatga ega. POS teglash usullarini rivojlantirish, xususan, Universal Dependencies kabi xalqaro andozalardan foydalanish ravishlarni to‘g‘ri belgilash va tushunish imkoniyatini oshiradi. Kelajakda o‘zbek tilida ravishlarning morfologik va semantik xususiyatlarini chuqurroq o‘rganish, ularni avtomatik tahlil qiluvchi texnologiyalarni takomillashtirish bo‘yicha izlanishlar davom etishi lozim.

Foydalanilgan adabiyotlar

1. G‘ulomov A. (2019). O‘zbek tilining grammatikasi. Toshkent, 45-bet.
2. Manning C. & Schütze H. (1999). Foundations of Statistical Natural Language Processing. MIT Press, 112-bet.
3. Xoliyorov U. (2020). O‘zbek tili morfologiyasi: Nazariy asoslar. Toshkent, 53-bet.
4. Salomov N. (2017). O‘zbek tilining stilistikasi. Toshkent, 27-bet.
5. Rasulov U. (2020). Zamonaviy o‘zbek tilshunosligi. Toshkent, 88-bet.
6. Jo‘rayev S. (2018). O‘zbek tili leksikologiyasi va morfologiyasi. Toshkent, 65-bet.
7. Jurafsky D. & Martin J. (2020). Speech and Language Processing. Prentice Hall, 86-bet

ARTIFICIAL INTELLIGENCE AND ONTOLOGICAL DICTIONARIES: MODERN APPROACHES

Abdullayeva Shaxnoza Raximjon qizi

Uzbekistan state world languages university 1st year PhD

E-mail: shakhnozaa40@gmail.com

Abstract: Ontological dictionaries are essential for knowledge representation and natural language processing (NLP) because they offer organized frameworks that connect machine-understandable concepts with human language. Significant progress has been made in the development, integration, and use of these dictionaries with the advent of artificial intelligence (AI), especially machine learning and deep learning methodologies. This study compares data-driven paradigms with traditional logic-based methodologies to examine modern techniques for creating and implementing ontological dictionaries in AI systems. There are case studies from biomedical text mining, intelligent assistants, and semantic search. The study ends with some thoughts on how multilingual ontologies, neuro-symbolic AI, and ontology learning might all intersect in the future.

Keywords: Artificial Intelligence, Ontological Dictionaries, Natural Language Processing, Knowledge Representation, Semantic Web, Ontology Learning, Deep Learning, NLP

Introduction

More advanced language tools that can describe knowledge in a fashion that is machine-interpretable have become necessary due to the rapid growth of artificial intelligence (AI). Semantic search, question-answering systems, and machine translation are just a few of the NLP tasks that rely on ontological dictionaries. These dictionaries map the relationships and characteristics of words, sentences, and concepts in an organized manner (ontology). This study looks at how AI and ontological dictionaries might work together, as well as contemporary methods for developing, improving, and using them.

Literature Review

Classical Ontological Dictionaries

Taxonomies derived from manually selected lexical semantics are offered by conventional ontological dictionaries such as WordNet, EuroWordNet, and FrameNet. Although these ontologies were essential to early AI systems, their scalability and domain flexibility were constrained.

AI in Ontology Learning

Recently, ontology construction has been automated through the use of AI, namely unsupervised learning and natural language processing models. To extract ontological relations from unprocessed text, programs such as Text2Onto and DL-Learner employ statistical techniques and logic reasoning.

Integration with Semantic Web Technologies

The OWL (Web Ontology Language) and RDF (Resource Description Framework) standards enable ontological dictionaries to integrate with the Semantic Web. AI agents leverage these resources for intelligent reasoning across the linked data environment.

Methodological Approaches

Symbolic AI + Ontology Engineering

Logic inference and rule-based systems are the foundation of symbolic AI techniques. These methods produce ontological dictionaries with easily comprehensible structures that are helpful in the fields of law, medicine, and regulation.

Statistical AI and Deep Learning

Ontological information is represented using transformer-based designs (e.g., BERT, GPT) and embeddings in recent models. The knowledge bases of ConceptNet and OpenAI dynamically relate semantic ideas via contextual embeddings.

Comparative Overview	Symbolic Ontologies	AI-Based Ontologies
Interpretability	High	Medium
Scalability	Low	High
Domain Adaptability	Medium	High
Human Effort	High	Low

Applications of AI and Ontological Dictionaries

Semantic Search

By determining user intent instead of just using keywords, ontologies improve search relevance. To enhance search results, Google's Knowledge Graph, for instance, makes use of ontological frameworks.

Intelligent Virtual Assistants

Systems like Siri, Alexa, and Google Assistant use ontological knowledge graphs combined with AI algorithms to interpret and respond to queries.

Biomedical Text Mining

In health informatics, tools like BioPortal integrate ontologies such as MeSH and SNOMED-CT to extract information from scientific literature using AI-based entity recognition models.

Challenges and Future Directions

Ontology Alignment and Merging

Terminologies may be inconsistent between ontologies. Although AI aids in the alignment of various vocabularies, maintaining semantic coherence is still difficult.

Multilingual and Cross-Cultural Ontologies

Developing multilingual ontologies that take cultural nuances into consideration is essential as AI grows more widespread. Promising paths for automated multilingual ontology generation are provided by AI models such as mBERT.

Neuro-symbolic Integration

Combining symbolic knowledge (ontologies) with neural architectures leads to neuro-symbolic AI—a promising approach that balances interpretability and learning capability.

Conclusion

From manually selected taxonomies to sophisticated, AI-enhanced frameworks, ontological dictionaries have seen significant development. Better semantic comprehension, interoperability, and cross-domain adaptability are made possible by their incorporation into contemporary AI systems. AI provides the tools required for the automatic creation, improvement, and use of ontologies as their size and complexity increase. The combination of organized ontological reasoning and deep learning suggests that AI systems will become more resilient, explicable, and context-aware in the future.

References:

1. Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*.
2. Navigli, R., & Velardi, P. (2004). Learning domain ontologies from document warehouses and dedicated websites. *Computational Linguistics*.
3. Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*.

4. Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web Journal*.
5. Camacho-Collados, J., & Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research*.
6. Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data – The story so far. *International Journal on Semantic Web and Information Systems*.

KORPUSGA ASOSLANGAN TURIZMGA OID MATNLARNING O'RGANILISHI

Adizova Nilufar IstamovnaBuxoro davlat universiteti O'zbek tilshunosligi
va jurnalistika kafedrası dotsenti, f.f.f.d.(PhD)E-mail: nilufaradizova8@gmail.com

Annotatsiya: Turizm dunyo hamjamiyatida mamlakatlar iqtisodiyoti, milliy madaniyati, tarixi hamda ijtimoiy hayotining ajralmas qismi, insoniyat faoliyatining barcha yo'nalishlariga bevosita daxldor, eng tez rivojlanayotgan, daromadli sohalaridan biridir. Ushbu maqolada turizm sohasining barcha turlarida keng qo'llanadigan madaniy va milliy jihatdan o'ziga xos konseptlarni nomlovchi, tasvirlovchi ifoda birliklariga xos ayrim xususiyatlar hamda ularning turistik matnlar asosida tilshunoslik nuqtayi nazardan dunyo olimlari tomonidan o'rganilishiga doir fikrlar o'z ifodasini topgan.

Kalit so'zlar: turistik matnlar, turistik birliklar, konsept, madaniy obyektlar, turizm, sayohat, umummilliy, hududiy, mintaqaviy birliklar.

Аннотация: Туризм является неотъемлемой частью экономики стран, национальной культуры, истории и социальной жизни в мировом сообществе, одной из наиболее быстро развивающихся и прибыльных сфер, непосредственно связанной со всеми направлениями человеческой деятельности. В данной статье освещены некоторые особенности выражений, обозначающих и описывающих культурные и национальные концепты, широко используемые во всех видах туристической деятельности, а также рассмотрены взгляды мировых учёных на изучение этих выражений в туристических текстах с лингвистической точки зрения.

Ключевые слова: туристические тексты, туристические единицы, концепт, культурные объекты, туризм, путешествие, общенациональные, территориальные, региональные единицы.

Annotation: Tourism is an integral part of a country's economy, national culture, history, and social life within the global community. It is one of the fastest-growing and most profitable sectors, directly connected to all areas of human activity. This article highlights certain features of expressions that denote and describe cultural and national concepts widely used across various types of tourism activities. It also examines the perspectives of international scholars on the study of these expressions in tourism-related texts from a linguistic point of view.

Keywords: tourism texts, tourism units, concept, cultural objects, tourism, travel, nationwide, territorial, regional units.

Jahon tilshunosligida turizmga oid matnlar va ularning lingvistik, leksik-semantik, diskursiv tadqiqiga doir korpusga asoslangan ilmiy izlanishlarga XX asrning ikkinchi yarmidan boshlab e'tibor kuchaydi. Sun'iy intellekt va turli tizimdagi axborot texnologiya vositalarining hayotimizdan keng o'rin ola boshlashi bugun bu kabi ishlarning yanada muhim ahamiyat kasb eta boshlaganini ko'rsatdi.

Dunyo tilshunosligida turizmga doir matnlar ekstralingvistik va lingvistik omillarning o'zaro ta'sirini aniqlash asosida o'rganilgan bo'lib, ularning asosiy qismi lingvokulturologik, antropotsentrik, pragmatik va kognitiv-diskursiv yo'nalishlarda amalga oshirilgan. Lingvistik jihatdan ularning leksik xususiyatlari esa, asosan, *terminologik, leksikografik va diskursiv* tomondan tahlil qilingan [Tayavi Yasen, 2017; 10]. Biroq o'zbek tilida turizm sohasiga oid leksik sath hamda matnlar tilshunoslik nuqtayi nazardan to'la asoslanib, korpus asosida tizimlashtirilmagan. Bu, ayniqsa, turizm sohasining barcha turlarida keng qo'llanadigan madaniy va milliy jihatdan o'ziga xos konseptlarni nomlovchi, tasvirlovchi ifoda birliklariga xos. Bu kabi

elementlar turizm sohasiga bevosita va bilvosita aloqador bo'lib, muayyan millatga xos milliy-madaniy tushunchalarni ifodalaydi hamda tilning leksik tarkibini boyitish va kengaytirishga jiddiy ta'sir ko'rsatadi.

L.M.Goncharova fikricha, “turizm” tushunchasi muayyan til sohiblari ongida mavjud bo'lgan “sayohat” lingvomadaniy konsepti bilan uzviy bog'liq holda predmetlarga emas, balki muayyan voqealarga oid murakkab mental xususiyatli konseptdir. Kontsept – bu har qanday madaniyat uchun muhim va qadrlil bo'lgan, u yoki bu obyekt haqidagi bilimlarni o'z ichiga olgan va tilda uni nomlovchi va tasvirlovchi leksik birliklar to'plamiga asoslangan milliy xususiyatga ega bo'lgan mental ta'limotdir[Goncharova L.M.,2012;58]. U muayyan bir millatning madaniy xotirasiga asoslangan bo'lib, keng semantik maydonga ega bo'lgan irod mahsullari(fenomeni) dir.

Bugungi kunda tarixiy va ijtimoiy shart-sharoitlar tufayli turizm sohasida ikki qarama-qarshi – turizm faoliyatining xalqaro hamda milliy lashuv tendensiyasi amal qiladi. Ushbu tendensiyalar asosida turizm sohasining o'ziga xos leksik sathi ham shakllanadi. Jumladan, sayohat uchun mo'ljallangan yo'riqnomalar, kataloglar, reklama matnlari, broshyura va qo'llanmalarda dominantlik vazifasini bajaruvchi milliy-madaniy birliklar, asosan, milliy lashuv tendensiyasi asosida shakllangan bo'lib, ular umummilliy, mintaqaviy, mahalliy hamda kasbiy[Tayavi Yasen, 2017; 10] xarakterga ega bo'ladi hamda axborot berish, ta'sir qilish, targ'ib qilish va kognitiv xususiyatlarni ifoda etadi.

Kasbiy ifoda birliklari, asosan, maxsus terminologiya sathi bilan bog'liq bo'lib, turizmga oid dunyoviy hamda sohaviy tushunchalar – terminlarni nomlash uchun qo'llanadi. Masalan, tur, turpaket, marshrut, ekskursiya, bron, recepsn(reception); sayohat agentligi, xona xizmati, safar obyektii kabilar.

Umummilliy konseptlar muayyan millatga xos madaniy jihatlarni aks ettirib, universal xarakter kasb etadi: “masjid”, “qal'a”, “imom”, “madrasa” kabi. Jumladan, pretsedent nomlar ham muayyan millat mentaliteti bilan bog'liq milliy-madaniy jihatlarni aks ettirish bilan bir qatorda universal ahamiyatga ham egadir. Masalan, Amir Temur, Bahouddin Naqshbandiy, Imom al-Buxoriy, Ibn Sino, Al-Xorazmiy, Abu Rayhon Beruniy, Jaloliddin Manguberdi va boshqalar. Chunki bu kabi birliklar allaqachon dunyoning aksariyat millat vakillari ongida muayyan assotsiatsiyalarni uyg'otib ulgurgan. Mintaqaviy birliklar ayrim hududda joylashgan madaniy merosni o'ziga xos tarzda nomlaydi, mahalliy ifoda birliklari esa muayyan bir joyning nomi bilan bog'liq voqelik asosida kodlanadi. Masalan, “Ko'kaldosh” nomi bilan ataluvchi masjid Buxoroda ham, Toshkentda ham bor yoki “yozgi saroy”, “xon”, “toq” kabi ifoda birliklar ham muayyan hududlarda o'ziga xos xususiyati bilan aks etadi. Mahalliy elementlar – “amir”, “Ark”, “Toqi sarrofon”, “Toqi zargaron”, “Chashmayi Ayyub”, “Somoniylar maqbarasi” kabilar faqat shu joy bilan bog'liq voqelikni yodga soladi.

Hozirda bu kabi milliy-madaniy jihatlariga, tarixiy-madaniy meros, osori atiqalarning nomlanishi, ularning mazmun-mohiyatini yoritib berish masalalariga e'tibor qaratish har qachongidan ham muhimdir. Shu sabab mamlakatimizda milliy istiqloldan keyin “bugungi global lashuv davrda har bir mustaqil davlat qatori o'z milliy manfaatini ta'minlash, bu borada, avvalo, o'z madaniyatini, azaliy qadriyatlarini, ona tilini asrab-avaylash va rivojlantirish masalasiga ustuvor ahamiyat qaratish”[O'zb. Res. PF-5850-сонли] muhim masalalardan biriga aylandi.

XX asrga kelib turizm jamiyatning juda tez rivojlanib borayotgan, integrallashgan, keng qamrovli sohalaridan biriga aylandi. Shu bois ushbu sohaga oid tarix(A.N.Dulov, K.A.Dyuxova, D.V.Yurchak), iqtisodiyot(R.Y.Stitsyuk; Tursunova G.R., Ochilova D.I.), madaniyatshunoslik(Moshnyaga Y.V., Dzyuba D.N., Vostrova Y.I., Vorobyova O.Y.), tarjimashunoslik(Marza N.E., Bepalova S.V., Kuznetsova L.N., Vasilyev S.K.), fizika-matematika(Qarshiyev A.B., Tursunov M.S.) kabi qator fanlar doirasida ko'plab, rang-barang tadqiqot ishlari yaratildi. Ayniqsa, turizm sohasiga doir matnlar lingvistik ishlarning muhim obyektii sifatida qator ilmiy ishlarning dunyoga kelishiga turtki bo'ldi.

Jahon tilshunosligida turizmga oid matnlarning til xususiyatlari monografik asosda T.I.Babkina, Y.A.Miroshnikova, O.P.Karenina, O.A.Kuzina, M.A.Kuzina, G.S.Atakyan, N.V.Filatova, E.Y.Novikova, M.G.Denisenko, A.S.Axmetshina, N.F.Starigina, V.A.Kolchevskaya, A.K.M.Alkudax, L.V.Vinogradova, M.Y.Plyushkina, D.Y.Gruzdev, T.Y.Zagryazkina, E.N.Bakurova, I.V.Odinsova, Qin Li, Shaobo Li, Sen Zhang Jie Hu, Zianjun Hu, O.V.Strijkova, I.V.Pak, S.N.Bezus kabi tadqiqotchilar tomonidan o'rganilgan bo'lib, ularning muhim qismi, asosan, matnlar tarjimasini bilan bog'liq lingvokulturologik, pragmatik, terminologik, reklama hamda diskursiv xususiyatlarini yoritish asosida amalga oshirilgan.

O'zbek tilshunosligida M.M. Buzrukova, S.S. Qodirov, N.T.Xodjayeva, N.K. Raimjanova, E.K. Elmuradova, Z.X. Umarova kabi olimlar tadqiqot ishlarini olib borishgan.

Korpus yondashuv asosida ham turizmga oid matnlarning diskursiv, pragmatik, qiyosiy tahlillariga ko'ra quyidagi natijalarga erishilgan: leksik to'plamlar, frazeologik birliklar aniqlanib, tahlil qilingan (Valensiya universiteti, Ispaniya), sayyohlarning Wuyuan qishloq turizmiga bo'lgan qiziqishlari, hududning tabiiy go'zalligi, an'anaviy arxitekturasini, tarixiy joylari va madaniy tajribalari bilan bog'liq afzalliklari web-matn materiallari asosida aniqlangan (Jinan Universiteti, Xitoy), Verona shahriga oid ingliz tilidagi ixtisoslashgan turistik reklama matnlaridan iborat korpusda modal fe'llarning kommunikativ funksiyalarining miqyosi va nozikliklari o'rganilgan (Italiya), vizual kontentni tez va keng auditoriyaga yetkazish uchun juda samarali platforma sifatida Instagramda sayyohlarni turli madaniyat, san'at, meros va urf-odatlar (madaniy turizm) bilan tanishtirishni targ'ib qilish va turli nutqlarda bitta WHS (Jahon merosi obyekti) haqida taqdim etilgan isbotlovchi turdagi ma'lumotlar (evidential information) va bilim asosidagi isbotlovchi ma'lumotlarning qanday aks etishi madaniyatlararo nuqtayi nazardan tahlil qilingan (Calabria Universiteti, Italiya)...

XX asrning 60-yillarida Germaniyada struktur-semantik yo'nalishning paydo bo'lishi bilan matnshunoslik alohida mustaqil fan sifatida shakllandi va rivojlandi. I. R. Galperin tilning matnga ta'sirini rad etmagan holda uni takrorlanmaydigan, "nutq ijodiy jarayonining ongli ravishda tashkil etilgan natijasi" sifatida til tizimiga emas, nutq tizimiga kiritdi. Matnga qiziqish tilshunoslikning nazariy rivojlanishi va matnni lingvistik birliklar tizimiga kiritishning tabiiy jarayoni, shuningdek, matnni tashkil etishning umumiy prinsiplari va ularni samarali qurish qoidalarini aniqlashdagi amaliy ehtiyojlardan kelib chiqqan bo'lib, fan va texnikaning tez rivojlanishi, xususan, kommunikatsiya nazariyasi, mashina tarjimasini, tillarni o'rgatish metodikasining takomillashuvi matnni turli fanlar nuqtayi nazaridan o'rganish va tahlil qilishga olib keldi: mantiq, sotsiologiya, psixologiya, ritorika, etnografiya, falsafa, san'atshunoslik va, albatta, lingvistika [Galperin I.P., 1981; 138].

Jumladan, lingvistik yondashuvlar asosida dunyo olimlari tomonidan olib borilgan ayrim ilmiy izlanishlarni kuzatadigan bo'lsak, ular, asosan, turizmga oid matnlarning quyidagi jihatlarini yoritib berishga ba'ishlanganini ko'rish mumkin. Masalan, D.Y.Gruzdev tadqiqot ishi turli tillarda milliy korpuslarni tuzish tajribasiga asoslangan bo'lib, ixtisoslashgan elektron matnlar korpuslari tuzish, ularni yozma tarjimon ishida qo'llash prinsiplari va usullarini aniqlash, shaxsiy tarzda tuzilgan tematik matnlar korpuslariga qo'yiladigan minimal talablarni belgilash, tarjimonlarning matnlar korpuslari va ularning munosabatlari hamda imkoniyatlari haqida xabardorligini oshirish omillarini o'rganishga bag'ishlangan [Gruzdev D.Y., 2013].

Y.M. Tayavining tadqiqot ishi fransuzcha sayohat qo'llanma va yo'riqnomalaridagi lingvokulturologik realiyalarni kompleks tarzda o'rganishga bag'ishlangan [Tayavi Yasen, 2017] bo'lib, muallif ularni turlarga ajratgan holda leksikografik, diskursiv xususiyatlari, eksplitsit va implitsit strategiyalari hamda professional va maxsus turistik terminlarning o'zaro farqli jihatlarini yoritishga diqqat qaratgan.

V.A.Kolchevskaya tahlillari asosida zamonaviy turistik internet-diskursning struktur-semantik, janr xususiyatlari tur operatorlar veb-sahifalari, turistik jurnallar, turistik ijtimoiy tarmoqlarning matnlari asosida o'rganilgan, tizimlashtirilgan va institutsional diskurslar tizimida persuzivlik (ishontirish) nutq ta'siri kategoriyasi sifatida uning lingvistik va nolingvistik vositalari

klassifikatsiyasi, pragmalingvistik, janr xususiyatlari, kommunikativ strategiya va taktikalari taqdim etilgan[Kolchevskaya V.A., 2020].

*N. Filatov*ning ishida turizm diskursining lingvistik xususiyatlarini aniqlash va tizimlashtirish, turizm sohasida bilimlarni kodlash va translyatsiya qilish usullari, kommunikativ strategiyalar va taktikalar, ularning til vositalarida ifodalanishi, turizm diskursining o'ziga xos lingvistik hamda tipologik xususiyatlarini tasvirlab berish, uni boshqa diskurslar orasidagi o'rnini aniqlash, uning perlokutiv salohiyati va manzilli auditoriyaga nutqiy ta'sirni optimallashtirishga yordam beruvchi til vositalarining xususiyatlarini o'rganish masalalari yoritib berilgan[Filatova N.V.,2014].

S.X.Lipiridi tomonidan 1200 dan ortiq onlayn lug'atlardagi maqola fragmentlari asosida turistik muloqot jarayonida qo'llangan 300 ta neologizm (rasmiy turistik leksika) tanlab olingan va onlayn-lug'at materiallari asosida turizm diskursining tarmoq inglizcha neologizmlari semantik-aksiologik va kognitiv-pragmatik jihatlari o'rganilgan; turizm leksikasining internetda yaratilishi, qayd etilishi va tarqatilishi jarayonida kognitiv jarayonlarning namoyon bo'lish xususiyatlari ko'rib chiqilgan hamda tarmoq madaniyatining sayyohlik kommunikatsiyasi ishtirokchilari til xotirasiga ta'siri, madaniy va sotsiolingvistik jihatlari tahlil qilingan[Lipiridi S.X., 2021].

*L.V.Vinogradov*ning ilmiy ishi [Vinogradova L.V.,2011] ingliz va rus tillaridagi turizm terminologiya tizimlarini sinxron va diaxron nuqtayi nazardan qiyoslab o'rganishga bag'ishlangan bo'lib, ingliz va rus tillaridagi turizm terminologiyasining muhim tipologik, struktur-semantik xususiyatlari hamda uning shakllanish jarayonlari kompleks tarzda o'rganilgan.

Novikova I.N. ishida "dam olish" lingvokulturologik kontsepti qadriyat komponenti sifatida birinchi marta chog'ishtirma tilshunoslikda ingliz, nemis va rus tillaridagi reklamalar shiorlarida universal qadriyat sifatida qo'llanishi, verballashishi hali lug'atlarda qayd etilmagan paremlar asosida semantik jihatdan o'rganilgan, tasniflangan, o'xshash va farqli jihatlari aniqlangan[Novikova E.Y.,2022].

Moshnyaga Y.V. tadqiqoti[Moshnyaga E.V.,2011] xalqaro turizm tizimidagi madaniyatlararo kommunikatsiya konsept sohasi, ijtimoiy-madaniy negizlari, uning paradigmali xususiyatlariga bag'ishlangan.

Goncharova L.M. turistik diskursni tahlil qilgan holda, "tushuncha" va "kontsept"ning farqli jihatlari ajratib ko'rsatadi hamda "sayohat"ni konsept sifatida "turizm"ni esa ma'lum bir inson faoliyati sohasiga tegishli tushuncha sifatida ko'rib chiqishni taklif qiladi[Goncharova L.M.,2012].

M.A.Kuzina tadqiqoti nemis tilidagi turizm bo'yicha matnlarda qo'llangan inglizcha so'zlarning nemis tilidagi leksik-semantik assimilyatsiyasiga bag'ishlangan bo'lib, ingliz va nemis tillarining asosiy xususiyatlari hisobga olingan holda ikki til xususiyatlari o'zaro taqqoslanib o'rganilgan[Kuzina M.A.,2006].

Plyushkina M. Y. dissertatsiyasi Rossiya va Britaniya matbuot reklamasidagi turizm xizmatlari uchun mo'ljallangan pretsedent fenomenlarni taqqoslab o'rganishga bag'ishlangan[Plyushkina M.Y.,2008].

Umuman olganda, kuzatishlarimiz shuni ko'rsatdiki, turizmga oid lingvistik tadqiqotlar, asosan, turistik matnlarni tarjima qilish masalalari, ularni reklama qilish yoki diskursiv xususiyatlarini yoritish doirasida hamda sohaga doir leksik birliklarning leksikografik tadqiqi asosida olib borilgan. Biroq ushbu sohaga oid ko'plab ishlar amalga oshirilgan bo'lsa-da, bugungi kunda mavjud ishlar turizmga oid matnlarning o'ziga xos xususiyatlari korpus tahlili asosida atroflicha qamrab olinmaganligini va bu muayyan tilda mavjud bo'lgan matnlarning milliy-madaniy xususiyatlari hamda sayohat doirasida qo'llanuvchi milliy-madaniy koloritni ifodalovchi leksik birliklar tahlili hamda korpus uchun matnlar to'plamini yaratishga ehtiyoj borligini ko'rsatadi. Bu esa yangi tadqiqot uchun asos bo'lib xizmat qiladi.

Foydalanilgan adabiyotlar

1. O'zbekiston Respublikasi Prezidentining "2017 – 2019-yillarda Buxoro shahri va Buxoro viloyatining turizm salohiyatini jadal rivojlantirish chora-tadbirlari to'g'risida" gi PQ-2980-son Farmoni\2017-yil, 19-may.
2. Беспалова С. В., Кузнецова Л. Н. Особенности жанра и композиции текстов немецкоязычных онлайн-путеводителей. Филологические науки. Вопросы теории и практики. 2023. Том 16. Выпуск 5.
3. Дулов А.Н., К.А.Дюхова, Д.В.Юрчак. Монография: История путешествий
4. и туризма-2011. 138 стр.
5. Дзюба Дарья Николаевна. Виртуальный музей в контексте цифровой культуры. Автореф. дисс. канд. культурологических наук – Саранск, 2019.
6. Elena Manca, Francesca Bianchi. Tourism and tourist promotion around the world: a linguistic and socio-cultural perspective. Salento-2013.
7. Филатова Наталья Вячеславовна. Дискурс сферы туризма в прагматическом и лингвистическом аспектах. Автореф. дисс. канд. филол. наук. - Москва, 2014.
8. Гальперин И. Р. Текст как объект лингвистического исследования. –М.: Наука, 1981. 138 с
9. Goncharova Л.М. Концепт «путешествие» и его межкультурная представленность в рекламной коммуникации сферы туризма. НИР. Современная коммуникативистика (№ 1, 2012). Москва.
10. Груздев Дмитрий Юрьевич. Электронный корпус текстов как эффективный инструмент переводчика. Автореф. дисс. канд. филол. наук - Москва-2013.
11. Istamovna, Adizova Nilufar. "Leksik birliklarning milliy o'zlik ruhini ifodalashdagi roli (Erkin Vohidov dostonlari asosida)." Buxoro: Soha lingvistikasi: tilshunoslik va tibbiyot integratsiyasining muammolari, yechimlari hamda istiqbollari (xalqaro konf.). Conferences. 2021.
12. Istamovna, A. N. (2024). Sayohat turizmiga oid matnlarda milliy-madaniy birliklarning leksik-pragmatik xususiyatlari. «contemporary technologies of computational linguistics», 2(22.04), 479-484.
13. Istamovna, A. N. (2023). Methodological possibility of intertextuality in artistic text. Innovation in the modern education system, 3(31), 182-191.
14. Istamovna, Adizova Nilufar. "Intermatnlar badiiy asar ta'sirchanligini oshiruvchi muhim vosita." journal of innovations in scientific and educational research 7.11 (2024): 209-217.
15. Колчевская Валерия Алексеевна. Средства выражения категории персуазивности в гипержанре «сайт» туристического интернет-дискурса. Автореф. дисс. канд. филол. наук - Краснодар – 2020
16. Кузина Марина Анатольевна. Особенности лексико-семантической ассимиляции английских заимствований в современном немецком языке (на материале текстов по туризму). Автореф. дисс. канд. фил. наук - Москва-2006.
17. Липириди София Христофоровна. Семантико-аксиологические и когнитивно прагматические особенности сетевых английских неологизмов туристического дискурса. Автореф. дисс. канд. филол. наук - Краснодар – 2021
18. Мошняга Елена Викторовна. Концептное пространство межкультурной коммуникации в системе международного туризма. Автореф. дисс. доктора. философских наук – Москва, 2011.
19. Nuria Edo Marzá. A Practical Corpus-based Approach to Teaching English for Tourism. International Journal of Applied Linguistics & English Literature. Vol. 3 No. 1; January-2014.
20. Новикова Э. Ю. Жанровые, ценностные и транслатологические характеристики международного туристического дискурса в немецкой и русской лингвокультурах. Автореф. дисс. канд. филол. наука. Уфа – 2019.С.13

21. Плюшкина Мария Юрьевна. Прецедентные феномены в российской и британской печатной рекламе услуг для туристов. Автореф. дисс. канд. филол. наук - Екатеринбург – 2008.
22. Qarshiyev Abduvali Berkinovich , Tursunov Muhammadsolih Sa'din o'g'li. O'zbek tili milliy korpusini loyihalash. "Kompyuter lingvistikasi: muammolar, yechim, istiqbollar" xalqaro ilmiy-amaliy konferensiya. Vol. 1 №. 01 (2022) B.94
23. Стыцук Р. Ю. Роль и место этнографического туризма в развитии российских регионов. Вестник Алтайской Академии экономики и права № 6, 2024. Tursunova G. R., Ochilova D. I. O'zbekistonda gastronomik turizmni rivojlantirishning Fransiya tajribasi. Samarqand-2024.
24. Тайави Я. М. Лингвокультурные реалии во французских туристических путеводителях. Дисс. канд. филол. наук. Воронеж, 2017. 25 стр.
25. Вострова Евгения Игоревна. Факторы и инструменты повышения туристической привлекательности регионов Российской Федерации. Автореф. дисс. канд. философских наук – Москва, 2023.
26. Воробьева Ольга Юрьевна. Музейные практики в малом провинциальном городе: углич в контексте ярославского региона . Автореф. дисс. канд. культурологических наук – Ярославль – 2022
27. Васильев Сергей Константинович. Лингвистический и культурологический анализ текста оригинала для перевода. Автореф. дисс. канд. филол. наук. - Москва, 2010.
28. Виноградова Людмила Валерьевна. Терминология туризма английского и русского языков в синхронном и диахронном аспектах. Автореф. дисс. канд. филол. наук - Великий Новгород — 2011.

INGLIZ VA O'ZBEK TILLARI FRAZEOLOGIK BIRLIKLARIDAGI ANTROPONIMIK KOMPONENTLARNING O'ZIGA XOS XUSUSIYATLARI

G'oyibov Umarjon Alisher o'g'li

O'zbekiston davlat jahon tillari universiteti tayanch doktorant

E-mail: iumar.goyibov@mail.ru

Annotatsiya: Frazeologik birlik bu – ikki yoki undan ortiq soʻzlarning birlashmasidan iborat, yagona semantik maʼnoga ega boʻlgan hamda boshqa leksik birliklardan oʻziga xos jihatlari bilan farq qiladigan turgʻun soʻz birikmalaridir, jumladan *throw the book at, the last straw, mukkası bilan tushib ketmoq, tesha tegmagan* kabilar. Frazeologik birliklar tarkibida antroponimik komponentlar shunday oʻziga xos jihatlarga ega boʻladiki, mazkur jihat ularni soʻz va soʻz birikmalaridan farqlashda hamda antroponimik komponentli frazeologik birliklarni ajratib olishda yordam beradi. Mazkur maqolada esa ingliz va oʻzbek tillarida antroponimik komponentli frazeologik birliklarning oʻziga xos jihatlari haqida soʻz yuritilib, ularga munosabat bildiriladi.

Kalit soʻzlar: frazeologik birlik, antroponimik komponent, oʻziga xos xususiyat, ijobiy va salbiy konnotatsiya, badiiy-ifodaviylik, ensiklopedik axborot.

Abstract: A phraseological unit is a combination of two or more words with a single semantic meaning, the outstanding feature that differs it from other lexical units, including *throw the book at, the last straw, mukkası bilan tushib ketmoq, tesha tegmagan*. In the structure of phraseological units, anthroponymic components have such specific aspects that it helps to distinguish them from a word and word combination, as well as to select phraseological units with an anthroponymic component. This article will discuss and overview the peculiarities of phraseological units with anthroponymic components in English and Uzbek.

Keywords: phraseological unit, anthroponymic component, peculiarity, positive and negative connotation, artistic-expressiveness, encyclopedic information.

Аннотация: Фразеологическая единица - это сочетание двух или более слов с единым семантическим значением, выдающаяся черта, отличающая ее от других лексических единиц, включая *throw the book at, the last straw, mukkası bilan tushib ketmoq, tesha tegmagan*. В структуре фразеологизмов антропонимические компоненты имеют такую специфику, что это помогает отличать их от слова и словосочетания, а также выделять фразеологизмы с антропонимическим компонентом. В данной статье будут рассмотрены особенности фразеологических единиц с антропонимическими компонентами в английском и узбекском языках.

Ключевые слова: фразеологическая единица, антропонимический компонент, своеобразие, положительная и отрицательная коннотация, художественно выразительность, энциклопедическая информация.

Ingliz va oʻzbek tillarining frazeologik birliklari tarkibida onomastik komponent sifatida ishtirok etgan antroponimlar ingliz va oʻzbek xalqi haqida ensiklopedik maʼlumot tashuvchi xususiyatga ega boʻlib, shuning asosida ularning alohida oʻziga xos xususiyatlarini ajratib olishimiz mumkin. Izlanishlarimiz natijasida ingliz va oʻzbek tillarida antroponimik komponentli frazeologik birliklarning oʻziga xos xususiyatlari sifatida quyidagilarni ajratib oldik: antroponimiklik, multikomponentlilik, strukturaviy xilma-xillik, semantik butunlik, turgʻunlik, takroriylik, variativlik, dinamiklik, majoziylik, pragmatiklik, kontekstlilik, ijobiy va salbiy konnotatsiyaga egalik, badiiy-ifodaviylik va muayyan xalqning qomusiy xazinası bilan bogʻliqlik.

Ingliz va oʻzbek tillari antroponimik komponentli frazeologik birliklarining antroponimikligi, shubhasiz, mazkur birliklarning koʻzga koʻringan birinchi xususiyati boʻlib, tarkibida bevosita kishining atoqli otlari ishtirok etishi bilan belgilanadi. Bunda mazkur birliklar

tarkibida bir yoki ikkita antroponimlar ishtirok etishi mumkin, masalan ingliz tilida bitta antroponimga ega frazeologik birliklar *As old as Methuselah*; *Hobson's choice*; *Jack Ketch* kabilarda aks etsa, ikkita antroponimga ega frazeologik birliklar *Rob Peter to pay Paul*; *Jack and Jill*; *David and Jonathan* kabilarda o'z ifodasini topadi. Bundan tashqari, ingliz tilida uchta antroponimga ega frazeologik birlik ham mavjud, masalan *Brown, Jones and Robinson* kabi. O'zbek tilida bitta antroponimga ega frazeologik birliklarni *Aflotun miya*; *Noz-noz, kal Niyoz*; *Majnun bo'lib qolmoq* singari misollarda, ikkita antroponimga ega frazeologik birliklarni esa *Eshmat ketib, Toshmat keldi*; *Alining alamini Validan olmoq*; *Aql – Hasan, odob – Husan* kabilarda kuzatishimiz mumkin.

Ingliz va o'zbek tillari antroponimik komponentli frazeologik birliklarining multikomponentliligi ularning tarkibida ikki komponentli so'zlar birikmasidan tortib to o'n va undan ortiq komponentli so'zlar birikmasining mavjudligi bilan belgilanadi. Masalan, sakkiz komponentli: ingliz tilida *Jack of all trades and master of none*, o'zbek tilida *Iskandar xo'ja, bulturg'jo'ja, bola ochadi, tashlab qochadi*.

Ingliz va o'zbek tillari antroponimik komponentli frazeologik birliklarining strukturaviy xilma-xilligi shu bilan belgilanadiki, ular bir emas, bir necha struktur-sintaktik tuzilishlardan iborat bo'ladi. Bunda ularning tarkibida qatnashgan antroponimlar qanday atoqli nomlardan foydalanilganligi, qaysi so'z turkumida kelishi, qanday grammatik vositalardan tarkib topganligi va nechta komponentdan iborat ekanligi ayon bo'ladi. Masalan, Ot + Ot struktur-sintaktik tuzilish asosida tarkib topgan frazeologik birliklar: ingliz tilida *Typhoid Mary*, o'zbek tilida *Aflotun miya*.

Ingliz va o'zbek tillari antroponimik komponentli frazeologik birliklarining semantik butunligi ularning, o'z tarkibidagi so'zlarning miqdoridan qat'iy nazar, aniq bir ma'noni anglatishi, muayyan bir semantik maydonni hosil qilishi bilan belgilanadi. Bunda muayyan ma'noni ifodalashda asosiy va muhim rolni mazkur birliklar tarkibidagi antroponimlar bajaradiki, ular bilan bog'liq ensiklopedik axborotidan kelib chiqib, muayyan bir shaxsning tashqi ko'rinishi, xulq-atvori, intellektual va ijobiy qobiliyatlari haqida ma'lumot beradi hamda muayyan bir obyektни tasvirlashda yoki ma'noni anglatishda muhim manba bo'lib xizmat qiladi. Bu esa, qo'shma so'zlar singari, mazkur birliklarning semantik tomoni ham murakkab va ko'pqirrali ma'no hodisasi ekanligini anglatadi [Tuxtayeva, 2004; 29]. Masalan, ingliz tilidagi *Cordelia's gift* frazeologik birligi ayol kishining muloyim va ingichka ovozi anglatib **Ovoz** semantik maydonini, *A banquet of Lucullus* frazeologik birligi mo'l-ko'l dasturxoniga ega katta bazm, shohona bazm ma'nosini bildirib, **Bazm** semantik maydonini hosil qiladi, va hokazo. O'zbek tilida esa *Alixo'ja-Xo'jaali* frazeologik birligi ikkala holatda ham bir xil mazmun va shaklga ega ma'nosini anglatib **Tenglik** semantik maydonini, *Bazmi Jamshid* frazeologik birligi hashamatli bazm, ko'rkam dasturxon va maishatbozlikni anglatib, **Bazm** semantik maydonini hosil qiladi, va hokazo.

Ingliz va o'zbek tillari antroponimik komponentli frazeologik birliklarining turg'unligi struktur jihatdan joylashuvi va ketma-ketligining barqarorligi hamda semantik jihatdan muayyan ma'nolarni anglatishda qat'iyligi bilan belgilanadi. Frazeologik birliklarning har bir komponenti shakl va mazmun birligidan iboratki, ushbu birliklarning ifoda va mazmun jihatlari so'zlarda bo'lgani kabi dialektik munosabatda bo'ladi, ya'ni biri ikkinchisiz mavjud bo'la olmaydi. Bunda frazeologik birliklarning ifoda plani ularning formal xususiyatlarini yuzaga chiqarsa, mazmun plani ma'no xususiyatlarini kashf qiladi [Abdusamadov, 2022].

Ingliz va o'zbek tillari antroponimik komponentli frazeologik birliklarining takroriyligi ularning og'zaki va yozma nutq jarayoniga qadar muayyan matnlarda va lug'at kitoblarida mavjudligi hamda ularda tez-tez va takroran qo'llanilishi bilan belgilanadi. Boshqacha so'z bilan aytganda tarkibida antroponim ishtirok etgan muayyan so'zlar birikmasi boshqa bir tarixiy voqea-hodisa, diniy va afsonaviy rivoyatlar, badiiy adabiyot durdonalari hamda xalq og'zaki ijodi namunalari allyuziv, ya'ni ishora sifatida ma'lum bir matnlarda, asarlarda yoki gazeta-jurnallarda qayta qo'llanilishi tufayli inson xotirasida tayyor frazeologik birlik sifatida yod bo'lib qoladi.

Ingliz va o'zbek tillari antroponimik komponentli frazeologik birliklarining variativligi ularning ma'nosi o'zgarmagan holda bir yoki bir nechta leksik-grammatik variantlari mavjudligi bilan belgilanadi. Masalan, ingliz tilida *the mark of Cain/ the brand of Cain* singari leksik variativlik, *the bed of Procrustes/ a Procrustean bed/ the Procrustes' bed* singari grammatik variativlik, o'zbek tilida esa *Abdulkhakim ovga chiqdi, orqasidan g'avg'o chiqdi/ Aliyor ovga chiqdi, ketidan g'avg'o chiqdi* kabi leksik variativlik, *Berdisini aytguncha urib o'ldirma/ Gapning Berdisini eshit/ Berdisini aytguncha urib o'ldirmoq/ Berdisini aytguncha urib o'ldiradigan/ Berdisini aytmoq* kabi grammatik variativlik hodisasini kuzatishimiz mumkin.

Ingliz va o'zbek tillari antroponimik komponentli frazeologik birliklarining dinamikligi muayyan maqsadlarda ularning tarkibidagi so'zlarning struktur-sintaktik jihatdan o'zgarishi bilan belgilanadi. Masalan, ingliz tilidagi *open a Pandora's box* antroponimik komponentli frazeologik birligi ko'plab chigal muammolarni keltirib chiqarmoq ma'nosini anglatib, u publitsistik nashrlarda, masalan "The New York Times" jurnalining 2013-yil 19-iyun sonida, *Pandora's box has been opened* kabi grammatik jihatdan hozirgi tugallangan zamon majhul nisbatiga o'zgartirilishi mumkin: *Pandora's box has been opened, and voices pleading for the deconstruction of the European project are becoming louder*. O'zbek tilida *Otasini so'rasang – Ahmadi forig', onasini so'rasang – tovonni yoriq* antroponimik komponentli frazeologik birligi haqiqatni biladigan lekin uni aytmaydigan yoki aytishga qo'rqadigan kishilarga nisbatan ishlatilib, u badiiy adabiyotda, masalan Tog'ay Murodning "Ot kishnagan oqshom" qissasida, *Otangni ko'rdim – Ahmadi forig', enangni ko'rdim – tovonni yoriq* kabi leksik-grammatik jihatdan substitutsiya hodisasiga o'zgartirilishi mumkin: *Otasini so'rasang – Ahmadi forig', onasini so'rasang – tovonni yoriq > Ulim, erta birov aytgan ekan, otangni ko'rdim – Ahmadi forig', enangni ko'rdim – tovonni yoriq, deb. Otang shul, enang shul. Qo'y endi, o'z tilingda gapir*.

Ingliz va o'zbek tillari antroponimik komponentli frazeologik birliklarining majoziyliigi ularning ko'pincha to'g'ri ma'noda emas, balki qisman yoki to'liq ko'chma ma'noda qo'llanilishi bilan belgilanadi. Bunda ushbu frazeologik birliklar ularning tarkibidagi antroponimlarga tegishli belgi-xususiyatlar inobatga olingan holda boshqa bir obyektни tasvirlashda qo'llanadi. Biz tasvirlamoqchi bo'lgan mazkur obyekt esa o'z ichiga bevosita insonni, shu jumladan, uning xarakteri, yurish-turishi, xatti-harakati va ruhiy kechinmalari, bundan tashqari, artefakt, oziq-ovqat, kiyim-kechak kabi borliqdagi muayyan predmetlar hamda abstraktsiya, ish-harakat, makon-zamon singari tushunchalarni qamrab oladi.

Ingliz va o'zbek tillari antroponimik komponentli frazeologik birliklarining pragmatikligi so'zlovchining turli-tuman hissiyotlarini tasvirlashi, tinglovchiga nisbatan aqliy, hissiy, xatti-harakatli ta'sir o'tkazishi bilan belgilanadi. Masalan, ingliz tilidagi *By George!* frazeologik birligi kishining ajablanish, hayratlanish, achchiqlanish, afsuslanish, ishonch, tilak kabi turli-tuman hissiyotlarini tasvirlab kelsa, *What will Mrs. Grundy say?* frazeologik birligi o'zgalar nima deb o'ylashlaridan yoki biror-bir fikr qonun-qoidalar va odob-axloq mezonlariga to'g'ri keladimi yoki yo'qmi deya xavotirlanishni bildiradi, shu bilan birgalikda, tinglovchiga qarata aytilgan mazkur frazeologik birlik yordamida so'zlovchining undan biror-bir maslahat yoki fikr-mushohada kutayotganligi ham anglashiladi. O'zbek tilida *Jabbor chalsin seni!* frazeologik birligi kishining achchiqlanishini, shu bilan birgalikda, biror-bir kimsa yoki narsaning ko'zdan yo'qolishiga istak bildirayotganligini ko'zda tutadi.

Ingliz va o'zbek tillari antroponimik komponentli frazeologik birliklarining kontekstlilik ularning ma'nolari nutq jarayonidagi muhit hamda shart-sharoitlarga bog'liq holda namoyon bo'lishi bilan belgilanadi. Ushbu kontekst muhiti esa frazeologik birliklarning ma'nosini shakllantiruvchi madaniy, ijtimoiy, tarixiy va lingvistik faktorlarni inobatga oladi [Pazilova & Sanginova, 2022]. Xususan, frazeologik birliklarning kontekstlilik belgisi zahirida muayyan turg'un birikmalarning ma'lum bir diskurs yoki kommunikativ kontekst muhitida qanday funksiyani bajarishini tadqiq qilish hamda mazkur turg'un birliklar ma'lum bir til va kommunikatsiya jarayonida qanday ma'noga, ifoda yo'siniga va ta'sirga egaligini aniqlash yotadi [Naciscione, 2010]. Masalan, ingliz tilidagi *Rob Peter to pay Paul* antroponimik komponentli

frazeologik birligi bir qarzdin qutulib, boshqasiga duchor bo‘lmoq, yomg‘irdan qutulib, qorga tutilmoq ma’nosini anglatadi. Mazkur frazeologik birlik siyosiy kontekstda ma’lum bir hukumatning bir byudjet mablag‘ini qoplash uchun boshqa bir byudjet mablag‘idan foydalanishni anglatadi, ijtimoiy kontekstda ma’lum bir kishining qarzini uzish maqsadida boshqa bir kishidan qarz olishni ifodalaydi. Badiiy kontekstda esa inson o‘z dunyoqarashining ma’lum bir nuqtayi-nazaridan boshqasiga o‘zgartirish kabi ma’nosini anglatishi mumkin. O‘zbek tilida *Sulaymon o‘ldi* – *devlar qutuldi* antroponimik komponentli frazeologik birligi biror yurt yoki oila boshlig‘i o‘z yurti yo oilasini qattiqqo‘llik bilan mahkam ushlab, tinchitib turgan, vaqti-soati yetib olamdan o‘tgach, yurt yomonlari yoxud noqobil farzandlar o‘z yomonliklarini qila boshlashini anglatadi. Mazkur frazeologik birlik siyosiy kontekstda o‘z qo‘l ostida hukumatni boshqarib turgan biror-bir yurt rahbari lavozimidan ketgach yoki olamdan o‘tgach o‘sha yurtning noqobil siyosiy arboblari o‘zboshimchalik bilan yurtni boshqarishini anglatadi, ijtimoiy kontekstda muayyan bir oilaning boshlig‘i biror yerga ketishi bilan farzandlarining sho‘xlik qilishi yoxud qaynonaning biror joyga mehmonga ketishi bilan kelinlarning uyda ish qilmay yotishini ko‘zda tutadi.

Ingliz va o‘zbek tillari antroponimik komponentli frazeologik birliklarining ijobiy va salbiy konnotatsiyaga egaligi – bu ularning ma’lum bir obyektini ijobiy va salbiy baholay olishi bilan belgilanadi. Bunda mazkur frazeologik birliklar tarkibidagi antroponimlar bilan bog‘liq qomusiy axborot asosida ularga tegishli belgi-xususiyatlar yo ijobiy yoki salbiy konnotatsiyaga egaligi namoyon bo‘ladi. Masalan, ijobiy konnotatsiyaga ega frazeologik birliklar: ingliz tilida *As patient as Job*, o‘zbek tilida *So‘zning Luqmon tilidan aytilgani* – *xo‘b*; salbiy konnotatsiyaga ega frazeologik birliklar: ingliz tilida *A doubting Thomas*, o‘zbek tilida *Bo‘ji bo‘ldi*, *bo‘ji keldi*, *Chingiz bilan Jo‘ji keldi*.

Ingliz va o‘zbek tillari antroponimik komponentli frazeologik birliklarining badiiy-ifodaviyligi – bu ularning muayyan manzarani obrazli tasvirlay olishi bilan belgilanadi. Bunda mazkur frazeologik birliklar tarkibida ishtirok etgan antroponimlarga xos belgi-xususiyatlar orqali biz borliq hodisalari va insonning ichki ruhiy kechinmalari singari muayyan bir manzarani obrazli tasvirlash imkoniga ega bo‘lamiz.

Ingliz va o‘zbek tillari antroponimik komponentli frazeologik birliklarining muayyan xalqning qomusiy xazinasi bilan bog‘liqligi – bu ularning muayyan millatga xos voqea-hodisa yoki kishi to‘g‘risida qomusiy axborot bera olishi bilan belgilanadi. Bunda mazkur frazeologik birliklar tarkibida ishtirok etgan antroponimlar muhim ahamiyatga ega bo‘lib, aynan ular orqali biz ingliz va o‘zbek xalqlari tarixi, madaniyati, adabiyoti, dini, yashash tarzi, folklori, afsona va rivoyatlarini o‘z ichiga olgan qomusiy xazinasi bilan, shu bilan birgalikda, ikki millatga xos mushtarak va farqli tomonlari bilan yaqindan tanishish imkoniga ega bo‘lamiz.

Xulosa o‘rnida shuni aytish mumkinki, ingliz va o‘zbek tillari antroponimik komponentli frazeologik birliklarining o‘ziga xos xususiyatlari mazkur frazeologik birliklarning shakllanishi, ma’nosi va qo‘llanilishida uning tarkibida ishtirok etgan antroponimik komponentlar muhim rol o‘ynaydi. Ushbu antroponimlar esa ingliz va o‘zbek xalqining tarixi, adabiyoti, dini, madaniyati, afsona va rivoyatlari, ijtimoiy hayoti haqida ma’lumot beradi.

Foydalanilgan adabiyotlar ro‘yxati:

1. Abdusamadov Z.N. Ingliz va o‘zbek tillarida antroponimik komponentli frazeologik birliklarning milliy-madaniy xususiyatlari: filol. fanl. falsafa dokt. (PhD) diss. – Toshkent: O‘zMU, 2022. – 132 b.
2. Qurbonova G.S. Fransuz va o‘zbek tillarida onomastik komponentli frazeologizmlarning milliy-lisoniy xususiyatlari: filol. fanl. falsafa dokt. (PhD) diss. avtoref.. – Toshkent: O‘zMU, 2019. – 50 b.
3. Radjabova M.A. Onomastik komponentli frazeologik birliklarning lingvomadaniy jihati va tarjima muammolari (ingliz, o‘zbek va rus tillari materialida): filol. fanl. falsafa dok. diss. avtoref. – Buxoro, 2021. – 51 b.
4. Tog‘ay Murod. Ot kishnagan oqshom: qissa. – Toshkent: Ziyo nashr, 2019. – 112 b.

5. Тухтаева К.Д. Структурные и семантико-номинативные характеристики сложных слов, входящих в состав фразеологизмов (на материале английского языка): Дисс. канд. филол. наук. – Навои: Навоийский ГПУ, 2004. – 162 с.
6. Naciscione A. Stylistic use of phraseological units in discourse. – Amsterdam: John Benjamins Publishing Company, 2010. – 292 p.
7. Pazilova N.M. & Sanginova D.A. Features of contextual meaning and translation issues of phraseological units in modern English and Uzbek // Oriental Renaissance: Innovative, Educational, Natural and Social Sciences. – 2022. – №4/2 (2). – P. 412–417.

HISTORICAL FOUNDATIONS AND EARLY FRAMEWORKS OF TEXT COMPLEXITY ASSESSMENT

Juraev Mukhammadrakhimkhon

Head of the department of Foreign languages
Jizzakh branch of National University of Uzbekistan named after M.U.

Annotation: the assessment of text complexity has undergone significant evolution, transitioning from early qualitative judgments to sophisticated multidimensional frameworks that shape contemporary educational practices. This study traces the historical trajectory of text complexity evaluation, beginning with foundational theories from Pestalozzi and Dewey, advancing through the development of quantitative readability formulas like Flesch-Kincaid and Lexile, and culminating in holistic approaches such as the College and Career Readiness Standards (CCRS). These paradigms reflect a growing recognition that text complexity encompasses measurable linguistic attributes, qualitative dimensions—such as thematic depth and coherence—and reader-specific factors, including background and motivation. Digital literacy further complicates this landscape, necessitating frameworks that address multimodal texts. By examining these shifts, this research underscores their profound impact on curriculum design, differentiated instruction, and literacy development, advocating for responsive pedagogical strategies that align texts with diverse student needs in an increasingly complex informational era.

Keywords: text complexity, readability formulas, literacy development, educational frameworks, qualitative assessment, quantitative measures, differentiated instruction, digital literacy, college and career readiness standards (CCRS), historical evolution

The evaluation of text complexity stands as a cornerstone of educational practice, directly informing the selection of instructional materials suited to diverse student populations. Understanding text complexity enables educators to enhance reading proficiency and cultivate enriched literacy environments, a pursuit that has evolved significantly over time. Historically, assessments of text complexity have progressed from subjective qualitative judgments to empirical quantitative measures, and more recently to multifaceted frameworks that integrate both approaches alongside reader and task considerations. Early educational theorists like Johann Heinrich Pestalozzi and John Dewey laid foundational principles by linking text accessibility to cognitive development and experiential learning, respectively. The 20th century marked a pivotal shift with the advent of readability formulas, such as Flesch-Kincaid and Gunning Fog, which provided systematic tools to gauge text difficulty. However, these methods faced criticism for their narrow focus on surface-level features, prompting subsequent frameworks to incorporate qualitative dimensions like thematic depth and coherence, as emphasized by scholars like Pearson and McKeown in the 1980s. The introduction of the College and Career Readiness Standards in 2010 further refined this approach, advocating a triadic model that balances quantitative metrics, qualitative analysis, and reader context. Today, the rise of digital literacy challenges traditional paradigms, demanding new methodologies to assess multimodal texts. This study explores the historical foundations and evolving frameworks of text complexity assessment, examining their implications for educational practices and literacy development, and highlighting the need for adaptive strategies in contemporary classrooms.

The evaluation of the complexity of the text is significant since it directly informs educators about the suitability of instructional materials for various student populations. By understanding and applying text complexity frames, educators aim to improve students reading capabilities and foster an enriched literacy environment. Historically, the evaluation of the complexity of the text has evolved through several frames and methodologies, ranging from early qualitative judgments to contemporary quantitative evaluations backed by sophisticated algorithms. This research seeks to explore the trajectory of the assessment of the complexity of the text, crucially examining its

impact on educational practices and literacy development. The introduction of various readability formulas, such as the Flesch-Kincaid and Lexile frameworks, represented a turning point, allowing educators to adopt empirical approaches to determine reading levels¹³². In contrast, the qualitative criteria established by frames such as the qualitative reading inventory underlined aspects such as thematic depth and organizational structure, which are essential for an integral understanding of the texts. These evolving paradigms reflect a broader recognition that the complexity of the text does not reside solely in measurable attributes, but is also influenced by the characteristics of the reader, which has significant implications for differentiated instruction. The current panorama of the complexity of the text is characterized by its multifaceted nature, as frames continue to expand the conceptualization of what interacts with complex texts.

The 1980s saw the introduction of more subtle structures that began to incorporate qualitative dimensions of textual analysis. Pearson and McKeown's work emphasized the need to consider factors beyond mere lexical and syntactic attributes in assessing the complexity of the text. His research highlighted the importance of thematic material, coherence, and intertextual connections of texts, postulating that these elements play a critical role in the way readers interact with texts. This change reflected an emerging understanding of reading as a complex cognitive process influenced by the reader's context, motivation, and strategic processing. The additional evolution in the evaluation parameters was evidenced by the adoption of Faculty and Career Readiness standards in 2010, which sought to outline a comprehensive text complexity structure through a multidimensional lens. The CCRs introduced a triadic approach to assess complexity, considering quantitative, qualitative, and reader/task considerations. This methodology allowed a more holistic assessment, recognizing that, while textual characteristics could be empirically measured, the effectiveness of reading and literacy development significantly depended on the alignment of the text with the interests, cultural background, and cognitive skills of the reader.

The impact of these evolutionary evaluation parameters reverberated through educational practices, causing changes in the curriculum design and pedagogical approaches. Educators are increasingly encouraged to adopt a responsive instructional posture, which allows different instructions, where text selection is aligned with the diverse needs of students. This responsive strategy is based on the understanding that the development of literacy is not only the involvement of an individual with the text, but is also improved through collaborative discussions, critical analysis, and scaffolding of understanding strategies.

In light of the rapid advances in technology, the emergence of digital literacy also requires a reassessment of text complexity measures. Current methodologies increasingly include the integration of multimedia texts and various formats, challenging traditional print paradigms. This transition emphasizes the dynamic interactions between text and reader, further complicating the evaluation of complexity but enriching the scenario of literacy practices.

As such, the historical trajectory of the complexity evaluation parameters of the text reveals a profound interconnection between educational structures and methodologies, illustrating its cumulative impact on the development of literacy. By critically analyzing these evolving parameters, stakeholders in education can better understand the dynamic intricacies at stake to promote effective reading practices that resonate with the various literate identities present in contemporary classrooms. The historical evolution of the evaluation of the complexity of the text can be traced back to the fundamental educational theories that have significantly influenced contemporary methodologies and practices. The first educational theorists, such as Johann Heinrich Pestalozzi and John Dewey in the 18th and 19th centuries, respectively, laid the foundations for a more nuanced understanding of literacy and their pedagogical implications. Pestalozzi, underlining the importance of natural learning processes, argued that the texts should be accessible to students, connecting the complexity of the text to cognitive development. In the meantime, Dewey's progressive education model continued this trajectory by supporting

¹³² Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233.

experiential learning, in which the complexity of the texts has become a critical factor in promoting critical thinking and commitment between students.

The 20th century marked a fundamental shift in the formalization of the evaluation of the complexity of the text. The introduction of readability formulas, such as the level of degree of Flesch-Kincaid and the Gunning Fog index, offered quantifiable measures to evaluate the difficulty of the text based on factors such as the length of the sentence and the familiarity of the words¹³³. These formulas have brought a more systematic approach to the selection of the text, guiding educators toward an empirical understanding of how the complexity of the text influences the results of understanding and learning. However, the addiction to mechanical assessments only has raised concerns regarding their validity in capturing the shaded dimensions of literacy, causing further evolution in evaluation practices. The main contribution of Flesch-Kincaid tests lies in their accessibility; Educators can quickly evaluate texts without specialized training in linguistics or literacy theory. This facilitated the pragmatic integration of text evaluations within educational environments, which allowed the selection of more properly challenging reading materials. However, this direct application also exposes the limitations inherent to such heuristic approaches. Bradley (2006) describes how the Flesch-Kincaid framework, while pioneering, mainly emphasizes the characteristics at the surface level of the texts, specifically the length of the sentence and the complexity of the words, at the expense of deeper textual characteristics, such as the thematic structure, the context and the implicit meanings that are crucial for the integral development of literacy¹³⁴. In addition, Flesch-Kincaid evaluations tend to undervalue the cognitive processes involved in reading comprehension¹³⁵. By selecting texts for students, the mere quantification of complexity based on measurable attributes does not explain the various histories of students, previous knowledge, and reading strategies, factors that significantly affect their interaction with texts. This excessive pedagogical simplification could lead to wrong instruction practices since educators can inadvertently favor texts that obtain adequate scores according to readability evaluations without considering their educational relevance or possible levels of participation for students.

CFER has also caused the development of several tools and evaluation resources that are explicitly focused on the complexity of the text. For example, the development of classification rubrics and performance descriptors associated with CEF levels allows educators to make more informed decisions about text selection. These tools allow a systematic approach to evaluate the complexity of the text based on established linguistic criteria, promoting consistency in educational evaluations. Organizations such as the British Council and Cambridge Evaluations have produced materials that operationalize CEF principles, thus extending their influence on classroom practices.

In addition, the impact of CFR reaches beyond European borders, since its principles and framework have been adapted by language programs around the world. Non-European contexts have increasingly recognized CEFR as a model to develop localized competition standards, which reflects the versatility of the framework to address various pedagogical needs and learning environments. As a result, discussions about the complexity of the text now incorporate global perspectives, recognizing the interaction between language learning and cultural literacy. In summary, CFR has fundamentally reformed the evaluation of the complexity of the text through its comprehensive framework, contextual emphasis, methodological advances, and global applicability. Its influence on educational practices and literacy development continues to resonate in several linguistic contexts, illustrating the essential role of the framework in modern language education. Within the domain of text complexity assessment, qualitatively directed methodologies have emerged as fundamental to contextualizing how texts are evaluated by their inherent

¹³³ Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233.

¹³⁴ Bradley, P. (2006). The history of simulation in medical education and possible future directions. *Medical education*, 40(3), 254–262.

¹³⁵ Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233.

complexities beyond quantifiable metrics. Hays and Singh (2011) elucidate the need to integrate qualitative research in educational environments, postulating that these structures can enrich the understanding of the complexity of the text, focusing on different interactions between readers and texts¹³⁶.

Historical structures that govern the assessment of the complexity of the text-as the qualitative, quantitative, and task considerations of the readers of the Central State standards, convenient a basis on which formative assessment strategies can be constructed. By substantiating their formative practices on well-established structures, educators can ensure that they are not only measuring the development of literacy effectively, but also promoting significant learning experiences that respect the different evolution of the complexity of the text in contemporary educational contexts., The analysis of the evaluation parameters of the complexity of the text reveals a convergence of theoretical paintings that underlines the multidimensional nature of the complexity of the text. Historically, the educational parties have recognized that reading and alphabetization are not simply products of the characteristics of individual readers or textual attributes; Rather, they are modeled by a complex interaction of cognitive, linguistic and contextual factors. This recognition is evident in various paintings that emerged over time, each of which contributes with the prospects faded to the evaluation of the complexity of the text. Among these frameworks is the Lexile framework for reading, which emphasizes quantitative measures such as the frequency of words and the duration of the sentence, providing a systematic approach to classify the texts based on their complexity. While this picture has been decisive in making the complexity of the text accessible through standardized metrics, it also attracted criticism for its reductionist vision of literacy, in which the richness of the literary commitment can be neglected. On the contrary, the qualitative measures introduced by the CCSS, which require an assessment of the complexity of the text along size such as structure, language and thematic content, reflect an evolution understanding of reading as a more holistic process, thus allowing a more complete evaluation of literary works . Furthermore, the gradual release of the responsibility model maintains that educators must have the commitment of students with increasingly complex texts. This pedagogy illustrates the dynamic nature of the complexity of the text by supporting a collaborative learning environment in which educators play a fundamental role in facilitating access to demanding texts. This approach is further supported by the Vygotskian theories of the proximal development area, highlighting how the appropriate challenges in reading can significantly improve the development of literacy if combined with adequate didactic support. The incorporation of multidimensional evaluation tools reflects the convergence of these paintings. The assessments that exploit the metrics both quantitative and qualitative, as well as factors such as the motivation of readers and cultural relevance, can provide educators with a solid understanding of the complexity of the text. This multidimensional approach recognizes that alphabetization is not a linear process limited to the act of reading alone, but a socio -social effort that takes place in specific educational and cultural contexts. The implementation of multidimensional evaluation practices has significant implications for educational practices. Educators are increasingly called to align the education with the needs of different students, recognizing that the complexity of the text is not exclusively inherent in a text but also derives from the interaction between the reader and the text. This requires the development of differentiated educational strategies that promote both involvement and understanding, further emphasizing the interaction between the complexity of the text and pedagogy. In short, the evolution of the evaluation parameters of the complexity of the text reflects an expanding understanding of the factors that influence the development of literacy. While the field moves towards more integrated approaches that consider cognitive, linguistic and cultural dimensions, educators have the task of applying these intuitions to promote a more fair and effective reading experience. The implications for the practice are profound, as they ask for a critical examination of curricular materials, educational strategies and evaluation methods to ensure that all students

¹³⁶ Hays, D. G., & Singh, A. A. (2011). *Qualitative inquiry in clinical and educational settings*. Guilford Press.

are equipped to navigate in increasingly complex texts in a rapidly evolving information landscape. In considering future directions for research on the complexity assessment of the text, it is imperative to integrate ideas of the existing literature while attracting connections with ongoing educational reforms.

The historical evolution of the evaluation of the complexity of the text illustrates a central journey marked by the iterative refinement of executives and methodologies in response to changing paradigms of educational practices and the development of literacy. From the beginning of the 20th century, when academics like Thormiike (1917) pleaded for empirical readability measures, to the contemporary recognition of the complexity of text with multiple face substantial transformation. Initially, readability formulas dominated the landscape, mainly focusing on the syntactic and lexical dimensions of the texts. The readability formula in Dale-Chall (1948) and the level of Flesch-Kincaid (1975), for example, offered quantitative measures that guided educators in the selection of appropriate reading equipment for students. However, these first executives were often criticized for their dependence on excess on the indicators at the level of the surface, by neglecting critical elements such as the knowledge of the content, the background of readers and the socio-cultural context of reading practices (Chall, 2000; Snow, 2002). In response to these limitations, more nuanced methodologies have emerged. The gradual integration of qualitative assessments has recognized the importance of the engagement of readers and the cognitive requests imposed by the texts. The CCSS has introduced a more complete approach by incorporating quantitative measurements and qualitative factors, thus recognizing that the complexity of the text includes elements such as the characteristics of language, structure and thematic depth (CCSS, 2010). This change has reflected an understanding that literacy is not only a mechanical competence but is influenced by the interaction between the characteristics of the text and the attributes of the readers.

The historical evolution of text complexity assessment reveals a dynamic interplay between educational theory, methodological innovation, and pedagogical practice, culminating in a nuanced understanding of literacy as a multidimensional process. From the qualitative insights of early theorists to the quantitative rigor of readability formulas, and onward to the integrative frameworks of the 21st century, this trajectory reflects a broadening recognition that text complexity transcends linguistic metrics alone. It is shaped by qualitative attributes, such as structure and thematic richness, and contextual factors, including readers' backgrounds, motivations, and interactions with increasingly diverse text formats, notably in digital contexts. These developments have reverberated through educational systems, fostering responsive instructional approaches that prioritize differentiated learning and align texts with students' unique needs. As digital literacy continues to redefine reading practices, future assessments must adapt to multimodal challenges, ensuring that educators are equipped to scaffold students' engagement with complex texts. This study underscores the cumulative impact of these frameworks on literacy development, calling for ongoing critical analysis and refinement to support equitable and effective reading experiences in a rapidly evolving educational landscape. The advent of digital literacy has further complicated the evaluation of the complexity of the text. While technological progress promotes new reading and commitment methods, executives are being adapted to meet the challenges posed by multimodal texts and the requirements of digital platforms. For example, the framework proposed by the New Literacies Research Lab recognizes that contemporary literacy implies critical thinking, problem solving and the ability to sail in various media (Coiro et al., 2008).

Used literature:

1. Chall, J. S. (2000). The academic achievement challenge: What works in the classroom? Guilford Press. (Inferred from your mention of "Chall, 2000" regarding socio-cultural context.)

2. Coiro, J., Knobel, M., Lankshear, C., & Leu, D. J. (Eds.). (2008). Handbook of research on new literacies. Routledge. (Cited as “Coiro et al., 2008” for the New Literacies Research Lab framework.)
3. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
4. Common Core State Standards Initiative (CCSS). (2010). Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects. National Governors Association Center for Best Practices & Council of Chief State School Officers. <http://www.corestandards.org/ELA-Literacy/> (Cited as “CCSS, 2010” for the triadic model.)
5. Council of Europe. (2001). Common European framework of reference for languages: Learning, teaching, assessment. Cambridge University Press. (Inferred from your references to CEFR and its global influence.)
6. Dale, E., & Chall, J. S. (1948). A formula for predicting readability: Instructions. Educational Research Bulletin, 27(2), 37–54. (Cited as “Dale-Chall, 1948” for readability formulas.)
7. Flesch, R. (1948). A new readability yardstick. Journal of Applied Psychology, 32(3), 221–233.
8. Hays, D. G., & Singh, A. A. (2011). Qualitative inquiry in clinical and educational settings. Guilford Press.

O'ZBEK TILIDA KOD ALMASHINUVI

Jo'rayeva Marjona Norbekovna

Jurnalistika va o'zbek filologiyasi fakulteti 1-bosqich talabasi

E-mail: jmarjona277@gmail.com

Annotatsiya: Mazkur maqolada kod almashinuvi jarayoni va uning zamonaviy o'zbek tilida namoyon bo'lish shakllari tahlil qilinadi. Kod almashinuvining asosiy sabablari sifatida global tillarning ta'siri, texnologiya va ilmiy atamalarning o'zbekcha muqobillarining yetishmasligi, shuningdek, ijtimoiy trendlar va moslashuv ehtiyojlari ko'rsatilgan. Misollar asosida kod almashinuvi jarayoni texnologiya, ijtimoiy tarmoqlar va ta'lim sohalarida keng qo'llanilayotganligi namoyish etiladi. Ushbu hodisa til rivojlanishining tabiiy qismi bo'lsa-da, u o'zbek tilining sofligi va madaniy identifikatsiya uchun muammolar keltirib chiqarishi mumkin.

Kalit so'zlar: kod almashinuvi, o'zbek tili, ingliz tili, rus tili, til rivojlanishi, ijtimoiy moslashuv, texnologiya, madaniy identifikatsiya, globallashtirish.

Abstract: This article examines the phenomenon of code-switching and its manifestation in modern Uzbek language. It highlights the primary reasons for code-switching, including the influence of global languages, the lack of Uzbek equivalents for technological and scientific terms, and social trends and adaptation needs. Through examples, the article illustrates how code-switching is widely used in technology, social media, and education. While this phenomenon is a natural part of language development, it may pose challenges to the purity of the Uzbek language and cultural identity.

Keywords: code-switching, Uzbek language, English language, Russian language, language development, social adaptation, technology, cultural identity, globalization

Аннотация: В статье рассматривается феномен кодового переключения и его проявления в современном узбекском языке. Основными причинами кодового переключения названы влияние глобальных языков, недостаток узбекских эквивалентов для технологических и научных терминов, а также социальные тренды и необходимость адаптации. На примерах демонстрируется широкое использование кодового переключения в технологиях, социальных сетях и образовании. Несмотря на то, что этот процесс является естественной частью развития языка, он может создать проблемы для сохранения чистоты узбекского языка и культурной идентичности.

Ключевые слова: кодовое переключение, узбекский язык, английский язык, русский язык, развитие языка, социальная адаптация, технологии, культурная идентичность, глобализация.

Kod almashinuv – bu ijtimoiy kontekst yoki suhbat muhitiga qarab, bir lingvistik koddan ikkinchisiga o'tish jarayoni. Masalan, bir gap ichida yoki nutq davomida boshqa bir tildan so'z, ibora, gap qo'shilishi. Kod almashinuv usullari sotsiolingvistlar, ijtimoiy psixologlar va ilmiy tadqiqotchilar tomonidan o'rganiladi. [1]

Hozirgi rivojlanyotgan zamonaviy muhitda o'zbek tilida kod almashinuvi bu — o'zbek tilining boshqa tillar, ayniqsa ingliz tili bilan aralashib, texnologiya, dasturlash, ijtimoiy tarmoq, turli xil shou va kundalik muloqotda qo'llanilishi. Internet, ijtimoiy tarmoqlar va IT sohasining rivojlanishi natijasida ko'plab texnik va raqamli atamalar o'zbek tiliga kirib kelmoqda. Aynan ijtimoiy tarmoqlar yordamida hozirgi yoshlar turli til kodlarini o'zlashtirib, o'zbek tili bilan bir qatorda foydalanyapti. Buning natijasida kontentlardagi soflik buzilib, kod almashinuv hodisasi yaqqol namoyon bo'lib qoldi. Eng hayratlanarli tomonlaridan biri shuki, bu hech kimni ajablantirmadi. Balki, bizda bu hodisalar qadimdan beri mavjudligidir. Hammamizga ma'lumki, turkiy til taraqqiyoti davomida forsiy til bilan yonma-yon bo'lgan. Natijada bu ikki til

orasida aralashuv paydo bo'lishi tabiiy holat. Ya'ni turkiy tilda so'zlashuvchilarning xoh ataydan, xoh behosdan bo'lsin forsiy so'zlarni qo'shib foydalanish holatlari kuzatilgan. O'sha davr adabiyotlari buni isbotlaydi. Masalan, "Sayohatnoma" janri asoschisi Muqimiy ijodini olaylik. Muqimiy yashagan yillar mustamlakachilikning ilk davrlariga to'g'ri keladi. Lekin u o'z she'rida ruscha so'zdan foydalanganiga e'tibor berganmisiz:

*Qiyg'ir ko'targon uch yigit,
 Uxlatmadilar bir minit,
 Ham boshda tong otquncha it
 G'ingshib chiqib, vang-vang ekan.*

"Minit" so'zi ruscha so'z bo'lib, uning o'zbek tilidagi muqobil varianti – *daqiqqa*. Shunda shoir bu so'zdan faqatgina qofiyalarning mos tushishi uchun foydalangan bo'ladimi? Yo'q, balki rus tili kirib kelishi bilanoq ommaviylashib boshlagani va ba'zi rus tilidagi kodlar hamma uchun tushunarli bo'lgani sabablidir. Agar o'sha davr turli qatlam aholisi bu lingvistik kodni tushunmaganida shoir bu so'zdan foydalanmasdi. Sababi Muqimiy o'z davrining mashhur shoirlaridan biri edi va uning ijodini katta qiziqishlar bilan qarshi olingan. Faqatgina elita qatlam tushuna oladigan so'zlardan foydalanganida bu asarlar oddiy aholi orasida yetarlicha shuhrat topmas edi.

Nafaqat mumtoz, balki zamonaviy o'zbek adabiyotida ta'sirchanlikni oshirish maqsadida kod almashinuv hodisasidan mahoratli foydalanilgan asarlar yetarlicha. Erkin A'zamning "Bayramdan boshqa kunlar" asari qahramoni – Safura nutqi bilan yaqindan tanish bo'lgan shaxslar darrov gap nima haqida ketyotganini payqashdi. Asardagi voqealar mustaqillik yillarining dastlabki kunlarida bo'ladi. Buni asar qahramonlari tilida saqlanib qolgan rus tilidan bilib olish mumkin. To'g'ri asosiy gaplar va tushunchalar o'zbek tilida berilsa-da, uslubiy bo'yoq hosil qiluvchi iltifot, kechirim, xursandchilik yoki arazlash kabi holatlarga tegishli bo'lgan kichik birliklar rus tilida keltirilgan. Misol uchun:

1. "*Kako-oy golos!*" – Safura tilidan olingan bu birlik, Bakirning ovozigga nisbatan hayratlanish vazifasida ishlatilgan.

2. "*Bravo Chinnibek*" – bu holatda esa u o'z xursandligini ingliz tilidan namuna keltirgan holatda davom etyapti.

3. "*Nu i shto*" – hattoki u o'z loqaydligini ham o'zi uchun oson bo'gan usulda ifoda etdi.

Hozirgi zamonaviy muhitda ham biz kattalar va yoshlar nutqini rus tili kodlarisiz sof holatda uchratishimiz anchayin mushkul. Buning ustiga, globallashuv jarayoni davomida yoshlarimiz nutqida nafaqat rus tili balki, ingliz, fransuz, turk tili namunalarini ham uchratmoqdamiz. Bunday almashinuv til rivojining tabiiy jarayoni bo'lib, u zamonaviy ehtiyojlarga moslashishni ko'rsatadi.

Unda, nega odamlar kod almashinuv hodisasini amalga oshiradi?

Ba'zan kod almashinuvi holatlari ataylab qilinadi — masalan, qandaydir tadbirga moslashish uchun kiyim yoki aksessuar tanlaganimiz kabi. Bu odamlar bilan "moslashish" yoki o'zini shu muhitga tegishli his qilish uchun.

Ba'zi holatlarda esa, bu avtomatik ravishda yuz beradi. Masalan, siz kim bilandir gaplashayotganingizda o'zingiz sezmasdan boshqa til, lahja yoki ohangga o'tib ketishingiz mumkin. Buni inson reflekslari boshqara olmaydi.[2]

Odamlar kod almashinuvi hodisasini ifoda qilish yo'lini yanada osonlashtirish, zamonaviy texnologik va ilmiy sohalaridagi so'zlarning yetishmasligi, ijtimoiy ta'sir va turli trendlar, ko'p tillilik muhiti va qulaylik sabablariga ko'ra ham amalga oshiradilar. Masalan, ba'zi tushunchalarni boshqa tilda ifodalash osonroq yoki aniqroq bo'ladi. "*update qildim*" degan ibora "*yangiladim*" dan ko'ra tezroq va zamonaviyroq eshitisli mumkin. Bundan tashqari IT, ilm-fan, marketing kabi yangi sohalarida ko'p atamalar hali o'zbek tilida to'liq shakllanmagan. Shuning uchun "*server*", "*platforma*", "*database*" kabi so'zlar asl holicha ishlatiladi. Shuningdek, yoshlar va internet foydalanuvchilari ko'pincha global trendlarga ergashadi. Ular uchun tilga inglizcha yoki ruscha so'zlar qo'shish "zamonaviylik" yoki "bilimdonlik" belgisi sifatida qabul qilingan. Ko'p tilli

muhitda yashovchi yoki ishlovchi odamlar (masalan, mehnat migrantlari yoki chet elda o'qiyotgan talabalar) tabiiy ravishda kod almashinuvi qiladi. Ba'zida biz uchun ham boshqa tildan biror atama yoki ibora ishlatish anchagina oson. Masalan, "*file ochdim*", "*zavtrak qildingizmi?*" yoki "*kompyuter obnovleniya bo'ldi*" kabi gaplar. Kod almashinuvi – bu til aralashuvi emas, balki sharoitga, maqsadga va muhitga til orqali moslashishdir.

Butun dunyo lingvistikasida, xususan, Amerika Qo'shma Shtatlarida kod almashinuvi dastlab, ikkinchi tilni o'zlashtirish jarayonida ya'ni, ispan tilida so'zlashuvchi shaxslarning ingliz tiliga o'tishi yoki aksincha holatlarida o'rganilgan. Jon J.Gumperz kabi yetuk sotsiolingvistlar esa aynan bir til doirasida so'zlashuvchi jamiyat a'zolarining o'z ona tillarini boshqa baynalmilal tillar bilan almashtirishga undagan ijtimoiy va madaniy omillarga ko'proq qiziqish bildirganlar. [3]

Amerikalik tilshunoslar Benjamin Li Uorf va Edvard Sapirning fikricha, til o'z foydaluvchilarining dunyoqarashi va shaxsini shakllantirish qudratiga ega. Ular ilgari surgan nazariyalar asosida tilida ko'p kod almashinuv hodisasi yuz beruvchi odamlar mafkurasini ikki turga ajratamiz: korreksionistik va kontrastivistik. Bunda tilini korreksionistik mafkura negizida shakllantirgan inson Yevropalashgan qadriyat, tushuncha va urf-odatlariga asoslangan madaniy o'ziga xoslikni qabul qilishi kuzatiladi. Ya'ni so'zlovchi jahon tillaridan birini o'zlashtirib, sekin-asta jahon madaniyatiga moslashishi nazarda tutiladi. Kontrastivistik mafkura asosida shakllangan inson esa, ijtimoiy kontekst yoki suhbat muhitiga qarab, boshqa bir o'ziga xoslikni qabul qilishi kuzatiladi. Biroq bu yerda so'zlovchi vaziyatga qarab hukmron lingvistik kod bilan o'z tili o'rtasida tanlov amalga oshirib, ulardan biri orqali nutqini davom ettirishi nazarda tutiladi. [4]

Bundan kelib chiqadiki, til inson tafakkuri va dunyoqarashining asosiy shakllantiruvchisi. Inson atrof-muhitni o'z tilining tuzilmasi va imkoniyatlari orqali qabul qiladi va anglaydi. Yuqoridagi nazariya nuqtai nazaridan qargalanda, kod almashinuvi (ya'ni bir tildan ikkinchisiga o'tish) nafaqat til o'zgarishi, balki dunyoqarash, madaniy moslashuv va shaxsiy o'zlikning qayta shakllanishi sifatida ham ko'riladi. Ya'ni korreksionistik yondashuvga ega inson o'z tili va madaniy qadriyatlarini global tushunchalarga moslab qayta shakllantiradi. O'z tilidan ko'ra jahon tillari (ayniqsa ingliz, rus yoki fransuz tili) ta'siri kuchli bo'lgan kodlar orqali fikrlay boshlaydi va u asta-sekin G'arb madaniyati, qadriyatlari va dunyoqarashini qabul qiladi. Masalan, yosh avlod ingliz tilini o'zlashtirish jarayonida nafaqat yangi so'zlarni, balki boshqa madaniyatning muayyan namunalarini ham qabul qilishi mumkin. Kontrastivistik yondashuvga ega inson esa, o'zligini yo'qotmagan holda turli kodlar orasida erkin harakat qiladi, har xil til va madaniyatlar o'rtasida muvozanatni saqlaydi. U suhbatlashayotgan shaxsga, ijtimoiy vaziyatga qarab, til kodlarini tanlaydi: masalan, oilada ona tilida, ishda rus yoki ingliz tilida gapiradi. Ammo bu til o'zgarishi uning o'zligidan voz kechganini anglatmaydi balki boshqa vaziyatga moslashganidan dalolat beradi.

Hozirgi XXI asr o'zbek tilida kod almashinuv jarayoni faol ravishda kuzatilmoqda. Quyida bu hodisani real misollar yordamida 3 aspektida ko'rib chiqamiz:

1. Kod almashinuvi qanday shakllarda namoyon bo'lmoqda?

Bugungi o'zbek tilida kod almashinuv hodisasi asosan ingliz va rus tillari bilan aralashuv shaklida yuz beradi.

Eng ko'p uchraydigan shakllar:

Intra-sentensial (gap ichida): "U Zoom call'ga kirmadi", "Deadline ni kecha topshirgandik". Agar so'zlovchi ikkinchi tilida qandaydir so'zni bilmasa yoki esidan chiqargan bo'lsa, birinchi tilidan so'z qo'shib gapirganida yuz beradi.

Inter-sentensial (gaplararo): "Men yangi loyihani boshlayapman. It's a startup."

Tag / Extra-sentensial – gap boshida yoki oxirida, so'zlar yoki qo'shimcha iboralar orqali.

[5]

Bu turlar kod almashinuvi funksiyalariga xizmat qilishi mumkin, faqat ular gap ichidagi joylashuviga qarab farqlanadi.

2. Kod almashinuvi yuz berayotgan asosiy sohalari:

a) Texnologiya va IT sohalari;

“Bugun deployment qilishimiz kerak”.

“Frontend bilan backend’ni connect qildik”.

b) Ijtimoiy tarmoqlar;

Bloglar, Instagram, Telegram kanal postlarida inglizcha va ruscha soʻzlarni aralashgan holda koʻrishimiz mumkin.

Misol: *“Yangi video content chiqqan. Linkni bio’ga qo’ydim”.*

c) Ta’lim va biznes muhiti;

IELTS, startup, pitch, accelerator, marketing, strategy kabi atamalar oʻzbekcha gap ichida toʻliq shaklda ishlatiladi.

3. Kod almashinuvining siyosatga ta’siri:

Oʻzbekiston hukumati rasmiy hujjatlarda oʻzbek tilini davlat tili sifatida mustahkamlasada, amaliyotda koʻplab davlat tashkilotlarida inglizcha-ruscha soʻzlar ishlatilmoqda.

Bu holat “Oʻzbek tili tozaligi” muammosini kun tartibiga olib chiqmaydimi?

Hozirgi oʻzbek tilida kod almashinuvi keng koʻlamda, ayniqsa, shahar yoshlari, IT sohasi mutaxassislari, va internet foydalanuvchilari orasida faol kechmoqda. Bu jarayon tilning dinamik rivojlanishidan, global tillar (ingliz va rus tillari) bilan yaqin aloqasidan dalolat beradi.

Foydalanilgan adabiyotlar:

1. www.britannica.com Carlos D.Morrison “Code-switching”
2. www.study.com Code switching in the classroom. Marie van Romburgh instructor: Della McGuire
3. www.health.clevelandclinic.org What is code switching.
4. Erkin A’zam, “Bayramdan boshqa kunlar”
5. Muqimiy “Sayohatnoma”

TURLI TIZIMLI TILLAR DOIRASIDA ONTOLOGIYANING NAZARIY ASOSLARI

Abduraxmonova Nilufar

filologiya fanlari doktori, professor,
O'zbekiston Milliy universiteti

E-mail: nilufarabdurakhmonova4@gmail.com

Boboxo'jayeva Laziza

Kompyuter lingvistikasi mutaxassisligi magistranti,
O'zbekiston Milliy universiteti

E-mail: Laziakhtamovna23@gmail.com

Annotatsiya: Mazkur tezis turli tizimli tillar doirasida ontologiyaning nazariy asoslarini o'rganishga bag'ishlangan. Unda falsafiy ontologiyadan lingvistik ontologiyaga o'tish jarayoni, lingvistik ontologiyaning maqsadi, vazifalari va uni yaratishning nazariy prinsiplari tahlil qilinadi. Tadqiqotning asosiy maqsadi turli tizimli tillarda (xususan, ingliz va rus tillari misolida) shakllangan yirik leksik-semantik ontologiyalarning nazariy jihatlarini aniqlash va ularni qiyosiy tahlil qilishdan iborat.

Kalit so'zlar: ontologiya, lingvistik ontologiya, semantik tarmoq, WordNet, RuWordNet, leksik-semantik tizimlar.

Annotation: This article is devoted to the study of the theoretical foundations of ontology within various systemic languages. It analyzes the transition from philosophical ontology to linguistic ontology, the goals and objectives of linguistic ontology, as well as the theoretical principles of its creation. The main aim of the research is to identify the theoretical aspects of large-scale lexical-semantic ontologies formed in various systemic languages (in particular, based on the examples of English and Russian) and to carry out their comparative analysis.

Keywords: ontology, linguistic ontology, semantic network, WordNet, RuWordNet, lexical-semantic systems.

Аннотация: Данная статья посвящена изучению теоретических основ онтологии в рамках различных системных языков. В нём анализируется переход от философской онтологии к лингвистической онтологии, цели и задачи лингвистической онтологии, а также теоретические принципы её построения. Основная цель исследования заключается в выявлении теоретических аспектов крупных лексико-семантических онтологий, сформированных на различных системных языках (в частности, на примере английского и русского языков), и их сравнительном анализе.

Ключевые слова: онтология, лингвистическая онтология, семантическая сеть, WordNet, RuWordNet, лексико-семантические системы.

Zamonaviy ilmiy tadqiqotlarning muhim yo'nalishlaridan biri bo'lgan ontologiya, borliqning fundamental tuzilishi, mavjudotlarning kategoriyalari va ularning o'zaro munosabatlarini o'rganadi. Falsafiy ildizlarga ega bo'lgan ushbu tushuncha XX asrning ikkinchi yarmidan boshlab lingvistika, kognitiv fanlar va sun'iy intellekt kabi sohalarda tobora muhim ahamiyat kasb etib kelmoqda. Ayniqsa, tilning voqelikni aks ettirish mexanizmlarini tushunish, bilimlar bazalarini yaratish va semantik veb texnologiyalarini rivojlantirishda lingvistik ontologiyalarning o'rne beqiyosdir [Gruber,1995].

Ontologiya atamasi falsafada borliqning eng umumiy tamoyillari, mavjudlikning asosiy kategoriyalari va ularning o'zaro munosabatlarini o'rganuvchi ta'limotni anglatadi. Falsafiy ontologiya "nima mavjud?", "mavjudlikning asosiy turlari qanday?", "ular qanday bog'langan?" kabi fundamental savollarga javob izlaydi.

Lingvistik ontologiya esa tilning leksik va grammatik tuzilishida voqelik qanday aks etishini o'rganuvchi soha bo'lib, til birliklari orqali ifodalangan kontseptual tuzilmalarni aniqlash va formalizatsiya qilishni maqsad qiladi. Bu soha tilning semantik tizimi orqali inson bilimlarining qanday shakllanishini tahlil qiladi.

Semantik tarmoqlar va leksik-semantik tizimlar lingvistik ontologiyalarning muhim tarkibiy qismlaridir. Semantik tarmoqlar kontseptlar va ularning munosabatlarini grafik ko'rinishda ifodalasa, leksik-semantik tizimlar esa so'zlarning ma'nolari va ular o'rtasidagi semantik aloqalarni (sinonimiya, antonimiya, giperonimiya va boshqalar) tizimli tarzda tashkil etadi. Ontologiyalar esa ushbu tizimlarni formal va aniq ifodalashga yordam beradi, bu esa bilimlar bazalarini yaratish va mantiqiy xulosalar chiqarish jarayonida muhim rol o'ynaydi [Fellbaum, 1998].

Ontologiyalarni yaratishda kontseptualizatsiya, taksonomik ierarxiya, semantik munosabatlarni aniqlash, formal ifodalash va baholash kabi nazariy prinsiplar qo'llaniladi. Yuqori darajadagi ontologiyalar (SUMO, DOLCE) eng umumiy tushunchalarni qamrab oladi, maxsuslashtirilgan ontologiyalar esa muayyan sohalarga oid bilimlarni modellashtirishga xizmat qiladi [Niles & Pease, 2001].

Turli tizimli tillardagi ontologiyalar (Ingliz tili).

Ingliz tilida semantik ma'lumotlarni tizimlashtirish va ontologik modellashtirish uchun bir qator yirik loyihalar amalga oshirilgan. Ulardan eng nufuzlisi WordNet hisoblanadi. WordNet – bu Prinston universiteti tomonidan ishlab chiqilgan leksik ma'lumotlar bazasi bo'lib, unda ingliz tilidagi so'zlar ma'nolariga ko'ra guruhlangan (sinsetlar) va turli semantik munosabatlar (sinonimiya, antonimiya, giperonimiya/hiponimiya, meronimiya/holonimiya va boshqalar) orqali o'zaro bog'langan. U so'zlarni kognitiv jihatdan yaqin bo'lgan ma'nolari bo'yicha guruhlashga intiladi [Miller, 1995].

WordNet uchta asosiy qismdan iborat: otlar, fe'llar, sifatlar va ravishlar. Har bir qism alohida taksonomik ierarxiya shaklida tashkil etilgan. Otlar ierarxiyasi “narsa” (entity) kabi umumiy tushunchadan boshlanib, tobora xususiy tushunchalarga qarab tarmoqlanadi. Fe'llar esa harakat turlari bo'yicha, sifatlar va ravishlar esa ularning semantik o'xshashligi va qarama-qarshiligi bo'yicha guruhlanadi. WordNet lingvistlar, kognitiv olimlar va tabiiy tilni qayta ishlash bo'yicha mutaxassislar uchun muhim resurs hisoblanadi. Uning nazariy asosi psixolingvistik tadqiqotlarga va inson leksik xotirasining tashkil etilishiga bog'liq.

WordNetdan tashqari, ingliz tilida SUMO (Suggested Upper Merged Ontology) kabi yuqori darajadagi ontologiyalar ham mavjud. SUMO turli sohalarga oid ontologiyalarni integratsiya qilish uchun umumiy ontologik ramka yaratishga qaratilgan. U IEEE Standard Upper Ontology Working Group tomonidan ishlab chiqilgan va keng qamrovli kontseptual ierarxiyani taqdim etadi.

Shuningdek, FrameNet loyihasi so'zlarning ma'nolarini ularning semantik freymlari (voqea yoki vaziyatni tasvirlovchi kontseptual tuzilmalar) orqali tushuntirishga intiladi. ConceptNet esa keng qamrovli commonsense bilimlar bazasi bo'lib, turli xil semantik munosabatlarni o'z ichiga oladi.

Turli tizimli tillardagi ontologiyalar (Rus tili).

Rus tilshunosligida ham ingliz tilidagi WordNetga o'xshash RuWordNet ontologiyasi mavjud. RuWordNet – bu rus tilidagi so'zlarning ma'nolarini sinsetlar orqali ifodalovchi va ular o'rtasidagi semantik munosabatlarni ko'rsatuvchi leksik-semantik baza. U Moskva davlat universiteti filologiya fakulteti tomonidan ishlab chiqilgan. RuWordNet tuzilishi jihatidan WordNetga o'xshash bo'lsa-da, rus tilining o'ziga xos leksik va semantik xususiyatlarini aks ettiradi. Masalan, rus tilining boy morfologiyasi sinsetlar va semantik munosabatlarning ifodalanishiga ta'sir qiladi [Azarova, 2016].

RuWordNetda ham otlar, fe'llar, sifatlar va ravishlar uchun alohida ierarxiyalar mavjud. Loyiha rus tilidagi so'zlarning ma'nolarini aniqlash, ularning sinonimik qatorlarini belgilash va turli semantik aloqalarini (masalan, “qism-butun”, “umumiy-xususiy”) ko'rsatishga qaratilgan.

Bundan tashqari, rus tilida PyTez kabi boshqa leksik-semantik resurslar ham mavjud bo'lib, ular turli semantik munosabatlarni (sinonimiya, antonimiya, analogiya va boshqalar) o'z ichiga oladi. WordNet.RU loyihasi ham rus tilidagi leksik-semantik aloqalarni tizimlashtirishga qaratilgan.

Rus tilining morfologik boyligi va so'z tartibidagi erkinligi kabi grammatik xususiyatlari rus tilidagi ontologiyalarning tuzilishi va semantik munosabatlarning ifodalanishiga o'ziga xos ta'sir ko'rsatishi mumkin.

Ingliz va rus tillaridagi ontologiyalarni qiyosiy tahlil qilish.

Ingliz va rus tillaridagi yirik ontologiyalarni qiyosiy tahlil qilish ularning tuzilishidagi o'xshashliklar va farqlarni aniqlashga imkon beradi. Masalan, WordNet va RuWordNet ikkalasi ham sinsetlar va semantik munosabatlarga asoslangan bo'lsa-da, ularning qamrovi, granularligi va semantik munosabatlarning aniqligi bo'yicha farqlar bo'lishi mumkin. Ikkala tildagi asosiy ontologiyalar (WordNet va RuWordNet) ham sinsetlar va semantik munosabatlar prinsipiga asoslangan. Biroq, ularning qamrovi, sinsetlar soni va semantik munosabatlarning aniqligi bo'yicha farqlar mavjud. Rus tilining morfologik boyligi RuWordNet tuzilishiga o'ziga xosliklar kiritishi mumkin [Журавлёв & Рогова, 2012].

Semantik munosabatlarning ifodalanishidagi xususiyatlar ham qiyosiy tahlilning muhim jihati hisoblanadi. Giperonimiya/hiponimiya, meronimiya/holonimiya, sinonimiya va antonimiya kabi asosiy semantik munosabatlar har ikki tilda ham mavjud, ammo ularning leksik vositalar orqali ifodalanishi tildan tilga farq qilishi mumkin. Masalan, rus tilida prefikslar va suffikslarning boyligi antonim va derivativ munosabatlarni ifodalashda muhim rol o'ynaydi.

Ontologiyalarning qamrovi, ya'ni qancha so'z va tushunchani o'z ichiga olishi hamda granularlik darajasi (tushunchalarning qanchalik batafsil tasvirlangani) tildan tilga farq qilishi mumkin. Bu farqlar tillarning leksik boyligi va ularda kontseptualizatsiya usullaridagi o'ziga xosliklar bilan bog'liq bo'lishi mumkin.

Madaniy va kognitiv farqlarning ontologiyalarda aks etishi ham qiziqarli tadqiqot yo'nalishi bo'lishi mumkin. Turli tillarda dunyoni kategoriyalash va tushunish usullaridagi farqlar ularning leksik-semantik tizimlarida va shunga mos ravishda ontologiyalarida ham o'z aksini topishi mumkin. WordNet asosan qo'lda leksikograflar tomonidan yaratilgan bo'lsa, RuWordNetni yaratishda ham qo'l mehnati katta o'rin tutadi, ammo avtomatik usullar va boshqa leksik resurslardan foydalanish darajasi farqlanadi.

Ontologiyalarni yaratish metodologiyalaridagi o'ziga xosliklar ham qiyosiy tahlilning muhim aspekti hisoblanadi. Turli tillardagi ontologiyalarni yaratishda qo'llanilgan prinsiplar, resurslar va avtomatlashtirish darajasida o'xshashlik hamda tafovutlar mavjud.

Xulosa qilib aytganda, turli tizimli tillar doirasida yaratilgan ontologiyalar, xususan ingliz va rus tillaridagi yirik leksik-semantik tizimlar, tilning semantik strukturalari va bilimlarni modellashtirish imkoniyatlarini chuqur tushunishga xizmat qiladi. Tadqiqot natijalari tillar va ontologiyalar o'rtasidagi o'zaro bog'liqlikni anglashga, hamda sun'iy intellekt, til texnologiyalari sohalarida samarali amaliy qo'llanmalarga asos yaratishga xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N., Qodirova Z. Kompyuter leksikografiyasida ontologiya (monografiya). Germany: Globe Edit, 2023.
2. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzworonet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
3. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus. uz. In 2021 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-4). IEEE.

4. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022, November). Developing NLP Tool for Linguistic Analysis of Turkic Languages. In 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 1790-1793). IEEE.
5. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
6. Azarova, I.A. "RuWordNet: Russian WordNet Development and Applications". Computational Linguistics and Intellectual Technologies, 2016.
7. Buitelaar, P., Cimiano, P., Magnini, B. "Ontology Learning from Text: Methods, Evaluation and Applications". IOS Press, 2005.
8. Fellbaum, C. "WordNet: An Electronic Lexical Database". MIT Press, 1998.
9. Gruber, T.R. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". International Journal of Human-Computer Studies, 1995.
10. Miller, G.A. "WordNet: A Lexical Database for English". Communications of the ACM, 1995.
11. Navigli, R. "Word Sense Disambiguation: A Survey"/ ACM Computing Surveys, 2009.
12. Niles, I., Pease, A. "Towards a Standard Upper Ontology" Proceedings of the International Conference on Formal Ontology in Information Systems, 2001.
13. Журавлёв, А.А., Рогова, А.Ю. "Онтологические модели в лингвистике". М.: Языки славянской культуры, 2012.

PARALLEL KORPUSNING O‘RNI VA AHAMIYATI

Abdujabbarova Nozima

O‘zbekiston Milliy Universiteti Jurnalistikava O‘zbek filologiyasi fakulteti
2-kurs 201-guruh talabasiE-mail: nozimaabdujabbarova2@gmail.com

Annotatsiya: Ushbu maqolada korpus lingvistikasi, korpus va uning bir yo‘nalishi bo‘lgan parallel korpus haqida fikrlar, uning tuzilishi va turlari haqida bayon etiladi. Bugungi kunda korpusning, ayniqsa, parallel korpusning o‘zbek tilshunosligi va tarjimashunoslikda elektron lug‘atlar yaratish, bir va bir nechta xorijiy tillarni o‘rganishdagi zamonaviy texnologiyalardan samarali foydalanish, avtomatik tarjima, xorijiy tillarni o‘rganishda tezkor qidiruv tizimi va kompyuter muammolari kabi bir qancha ishlar aynan parallel korpus orqali amalga oshirilishi nazarda tutiladi. Jahon hamjamiyatida, shuningdek o‘zbek tilshunosligida parallel korpus rivoji ham aynan ta’lim va texnologiya bilan chambarchas bog‘liq holatda o‘rganilishi ahamiyatlidir.

Kalit so‘zlar: korpus, parallel korpus, me‘zonlar, tadqiqot, ma’lumot, ilmiy ishlar

Abstract: This article discusses corpus linguistics, corpora, and one of its branches—parallel corpora—including their structure and types. Today, corpora, especially parallel corpora, are considered crucial in Uzbek linguistics and translation studies for tasks such as creating electronic dictionaries, learning one or multiple foreign languages using modern technologies, automatic translation, fast search systems in language learning, and addressing computational issues. The development of parallel corpora, both in the global community and in Uzbek linguistics, is notably important and is closely linked with education and technology.

Keywords: corpus, parallel corpus, criteria, research, data, scientific studies.

Аннотация: В данной статье рассматриваются корпусная лингвистика, корпус и одно из её направлений — параллельный корпус, а также их структура и виды. В настоящее время корпуса, особенно параллельные корпуса, играют важную роль в узбекском языкознании и переводе: они используются для создания электронных словарей, изучения одного или нескольких иностранных языков с применением современных технологий, автоматического перевода, быстрого поиска при изучении иностранных языков, а также для решения компьютерных задач. Развитие параллельных корпусов как в мировом сообществе, так и в узбекском языкознании является актуальным и тесно связано с образованием и технологиями.

Ключевые слова: корпус, параллельный корпус, критерии, исследование, данные, научные работы.

Kirish

Bugungi kunda dunyo miqiyosida korpus lingvistikasi asosan rivojlanib borishi orqali turli til va unga oid bo‘lgan qator muammolar o‘z yechimini topmoqda. Jumladan, korpus yaratish nimaga kerak? - degan savol o‘rtaga tashlansa, asosiy sabab sifatida tilning o‘lik tillar kabi yo‘q bo‘lib ketishini oldini olish va milliy tillarni saqlab qolish kabi ustuvor maqsad sifatida, so‘ng esa til o‘rganuvchilar, leksikolingvistlar, tarjima va turli til bilan aloqador bo‘lgan matnlarni qayta ishlashda korpus lingvistikasi asosiy mezondir. Hozirgi kunda jamiyatimizda texnologiya va texnika rivojlangan bir davrni o‘z ichiga olgan muhitda bu sohaning mavjudligi va rivoji muhim ahamiyatlidir. Bugungi kunda o‘zbek kompyuter lingvistikasi, tabiiy tilni qayta ishlash (NLP), mashinali ta’lim (machine learning), ma’lumot qidiruvi (Data Mining) kabi yangi sohalar kesimida rivojlanmoqda hamda ushbu sohalarda erishilayotgan ilmiy-amaliy natijalar bir-birini to‘ldirib bormoqda. O‘zbekiston Respublikasi Prezidentining 2019-yil 21-oktabrdagi “O‘zbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan oshirish chora-tadbirlari to‘g‘risida”gi PF-

5850-son farmoniga muvofiq davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalarining integratsiyalashuvini ta'minlashda qator ustuvor vazifalar ko'rsatib o'tildi.[Sh.Mirziyoyev,2020:PF-6084-son farmoni //https://lex.uz/docs/5058351.] Shu tariqa O'zbekistonda korpusga aloqador bo'lgan bir qancha olimlarning tadqiqotlari o'z o'rniga ega. Jumladan, o'zbek korpus lingvistikasida B.Mengliyev, Sh.Shahobiddinova, Z.Xolmanova, S.Karimov, L.Raupova, Sh.Hamroyeva, N.Abduraxmonova, G.Toirova, J.Djumabayeva, G.Ergasheva, A.Eshmominov, Sh.Gulyamovalarning tadqiqotlarini qayd etilgan.[B.Mengliyev, 2020:55.]

Muhokama va natijalar

Parallel matnlar korpusining dastlabki namunasi sifatida 1799-yilda Rozetta shahri yaqinidagi Nil deltasida topilgan mil.av 196-yilga borib taqaladigan, Misr ibodatxonalari tomonidan ikki tilda (yunon va misr) va uchta yozuv tizimida (Misr iyerogliflari va demotik tilda yozilgan) shoh Ptolomey Vga berilgan sharaflar haqida so'z yuritilgan tosh hisoblanadi. Dastlabki birinchi korpus 1897-98-yillarda nemis tilshunosi Keding tomonidan yaratilgan. U qog'oz shaklida bo'lib so'zlarda tovushlarning taqsimlanish chastotasini tekshirish maqsadida tuzilgan edi. Ammo katta hajmli ma'lumotlar qo'lda hisoblab chiqish mushkulligi uchun tadqiqotchilar konkordans keyin korpus menejerini ishlab chiqdi. Korpusga qo'yilgan talablardan biri uning interfeysga egaligidir. Shu asnoda korpus rivojlanib turlarga bo'lina boshladi.

Korpusning asosiy turlaridan biri bo'lgan parallel korpuslarning birinchi parallel matnlari Shveytsariyada nemis, fransuz, italyan tillarida to'plangan qor ko'chkisi haqidagi hisobotlar, Kanada ommaviy axborot vositalari ingliz, fransuz tillarida berilgan ob-havo ma'lumotlari 1980-yillarning oxiri 1990-yillarning boshida paydo bo'lgan. Dastlabki elektron resurslardan biri Kanada Hansard bo'lib, u dastlab jumalarni moslashtirish uchun ishlatilgan.(Gale and Church 1991), bu vazifa hozirda tarjima xotiralari kabi ilovalarni standart xususiyati hisoblanadi. Bundan tashqari, parallel korpuslar mashina tarjimasi, tilni ikkinchi til sifatida o'qitish hamda qiyosiy tilshunoslik sohalarida asosiy o'rganish obyektidir. Parallel korpus bir necha tildagi matnlar majmuasidan iborat bo'lib, ma'lumotlar lingvistik bazada tarjima muqobillari asosida o'zaro bog'lanadi. Masalan, rasmiy saytlardagi ma'lumot yoki yangiliklar ikki yoxud undan ortiq tillarda beriladi. Bu o'z navbatida matnning tarjima muqobillari tarzida ma'lumotlar bazasi (MB) sifatida yig'ilib mashina tarjimasi uchun resurs bo'lib xizmat qiladi. Multilingval (ko'p tilli) korpuslarning to'plami parallel korpuslardan tashkil topmay, berilgan matn boshqa tilda analogi bo'lmagan mazmun jihatdan bir mavzuga yo'naltirilgan matnlar majmuasidan iborat bo'ladi. Gap qurilishi va leksik sathi aynan o'xshash bo'lmasa-da, undan olinadigan kalit so'zlar statistikasi bir-biriga yaqin turadi. Endilikda parallel korpuslar turli tillar doirasida takomillashib bormoqda. Dunyo tajribasida korpus yaratishning lingvistik, matematik va dasturiy tomonlari olimlarining tadqiqotlarida o'z ifodasini topgan. Parallel korpusning tuzilishi, tarkibi va imkoniyati haqidagi ma'lumotlarni D. O. Dobrovolskiy, Yu. Tao, V. Zaxarov, A. A. Kokoreva, E. P. Sosninalarning ishlarida kuzatamiz. Anarbayev Orzubek Raxmanovich parallel korpuslar haqida shunday deydi. „Parallel korpus madaniyatlararo muloqot keng yoyilgan hozirgi davr uchun muhim voqelik hisoblanadi.Parallel korpuslar orqali turli til muhitidagi, madaniyatlaridagi, universalialar hamda tillarning o'ziga xos mental xususiyatlari, realiya va lakunar birliklarni aniqlash mumkin bo'ladi.Parallel matnlar korpusi avtomatik tarjima rivoji uchun xizmat qiladi, kompyuter leksikografiyasining taraqqiyotini ta'minlaydi.”

Parallel korpus turlari:

1)„Shovqinli“ parallel korpus—bu u darajada sifatli tarjimaga ega bo'lmagan hamda tarjima muqobillari ham mukammal darajada muvofiqlashmagan bo'ladi.

2)Qiyosiy parallel korpus—bu mavzu jihatdan mushtarak, biroq aynan matnning sintaktik va semantik jihatdan bir-biridan farqli korpus turiga kiradi.

Parallellik mezoniga ko'ra, korpuslar ikkita asosiy turga bolinadi:

1) haqiqiy parallel korpuslar (translation corpora) — muayyan tilda yozilgan ko‘plab asl matnlar va ushbu matnlarning bir yoki bir nechta boshqa tillarga tarjima qilingan matnlari mavjud korpuslar.

2) qiyosiy korpuslar (comparable corpora) — matnlari biron bir belgi asosida birlashtirilgan korpuslar. Bunday korpuslarda, masalan, o‘zbek va ingliz tillaridagi muayyan mavzu doirasidagi matnlar (bir xil tematik maydon, bitta tilning dialektlari, mintaqaviy variantlari, masalan, o‘zbek tili- ona tili va o‘zbek tili chet tili va boshqalar) o‘z aksini topadi. [Abjalova M., 2022:110.]

Liverpul universitetida 2009-yilgi Korpus tilshunosligi konferensiyasi bo‘lib o‘tadi, unda parallel korpuslar bilan ishlaydigan tilshunoslar va tarjimashunoslik talablari, korpusdan o‘z maqsadlari uchun foydalanish asboblari hamda korpus interfeysi bilan bog‘liq masalalar muhokama qilinadi.[<https://ares.uz>] Parallel matnlardan foydalanish uchun eng birinchi bo‘lib, matn segmentlarini (iboralar, yoki jumalarni) identifikatsiyalash va matn turlari va janrlarini o‘zaro moslashtirish, muvofiqlashtirishimiz kerak. Buni bosqichma-bosqich amalga oshirish tartibi mavjud. Matnlarni tahlil qiluvchi avtomatik analizatorni yaratishda, avvalo, elektron morfologik lug‘at yaratiladi. Morfologik tahlil natijasida so‘z turkumlariga oid statistik natija olinadi. Shu tarzda leksemalarning katta mantiqiy bazasi yuzaga keltirildi. Morfologik tahlil sintaktik tahlil uchun ham muhimdir. Tillarning leksik birliklarining semantikasini aniq ochib berishda parallel korpus ko‘prik vazifasini o‘taydi. Parallel korpuslarda matndagi kalit so‘zlarni ajratib olishimiz ishimizga yengillik beradi.

Parallel korpuslarga namunalar:

TAUS Data Cloud: <https://datamarketplace.taus.net/search?s=54&t=740>

«Слова о полку Игореве» <http://nevmenandr.net/slovo>

The JRC-Acquis Multilingual Parallel Corpus of the total body of European Union (EU) law: Acquis Communautaire with 231 language pairs.

EUROPARL — o‘n bir tildagi Yevropa parlamentidan olingan yigirma million so‘zli korpus (<http://www.isi.edu/~koehn/publications/europarl/>);

CHEMNITZ GERMAN-ENGLISH TRANSLATION CORPUS —bir million so‘zdan iborat bo‘lib, ilmiy, siyosiy va turizm sohasiga doir korpus (<http://www.tuchemnitz.de/phil/InternetGrammar>);

KACENKA — ingliz-chex tili parallel korpusida uch million so‘z mavjud (<http://www.phil.muni.cz/angl/kacanka/kachna.html>);

OPUS — ochiq resursli parallel korpus besh tildagi matnlar majmuasidan iborat. (<http://logos.uio.no/opus/>);

BROWN korpusida ingliz-franzus parallel korpusi o‘rin olgan;

INTERSECT — ingliz-nemis parallel korpusi shular jumlasiga kiradi.

Glosbe: <https://glosbe.com/>

Linguee: <https://www.linguee.ru/>

MyMemory: <https://mymemory.translated.net>

Reverso: <https://www.context.reverso.net>

Xulosa

Bugungi kunda korpus barcha uchun vaqt va samaradorlik kalitidir. Korpusga asoslangan til tarixi, korpusning morfologik va semantik analizatorligi, qidiruv tizimi va mualliflik holati, tabiiy tilni qayta ishlash va mashina tarjimasi kabi bosqichlari, xususan, o‘zbek korpusi uchun hamda jahon korpuslarida alohida ahamiyatga egadir. Korpus oqrali tilda so‘zlarning kengayishi torayishi, neologizmlar, sinonimiya va yangi fraziologiyalarning paydo bo‘lishini kuzatish mumkin. Parallel korpus orqali esa tarjimadagi g‘alazliklarga chek qo‘yish va tarjima sifatini oshirish ko‘zda tutilmoqda. Parallel korpuslardan kutilyotgan natija nafaqat tarjimonchilikda, balki mashina tarjimasida qo‘llashdir. Parallel korpuslar til korpuslarning maqsadiga va tillariga ko‘ra alohida turi bolib, uning bazasida ikki til yoxud bir tilning ikki xil xususiyatiga xos matnlar jamlangan bo‘ladi. Shu bois uning nomi ko‘plikda qo‘llaniladi, yani parallel korpus emas, parallel

korpuslar. Mazkur maqolada parallel korpuslarning kelib chiqishi va ulardan foydalanilish omillari, parallel korpuslarning ahamiyati borasida soʻz yuritildi.

Foydalanilgan adabiyotlar roʻyxati:

1. Abduraxmonova N. Oʻzbek tili elektron korpusining kompyuter modellari. Globedit, 2021. 220 b.
2. Abjalova M. Korpus lingvistikasi. [Matn]: uslubiy qoʻllanma / M.A. Abjalova. — Toshkent: Nodirabegim, 2022. — 110 b.
3. Mengliyev B., Hamrayeva Sh. Korpus lingvistikasi: korpus tuzish va undan foydalanish. — T.: Globeedit. 2020. — 50 b.
4. Захаров В. П., Богданова С.Ю. Корпусная лингвистика. Учебник для
5. студентов гуманитарных вузов. — Иркутск: ИГЛУ, 2011. — С. 16.
6. Kurdi M.Z. Natural Language Processing and Computational Linguistics:
7. Speech, Morphology and Syntax. — Great Britain, USA: Wiley-ISTE, 2016. — P. 4.
8. Каримуллина Р.Н., Каримуллина Г.Н. О сводном лингвографическом корпусе татарского языка / Слово и словарь (vocabulum et vocabularium)
9. Сборник научных материалов. МИНСК ИЗДАТЕЛЬСТВО «ЧЕТЫРЕ ЧЕТВЕРТИ» 2017. — С. 18.
10. <https://uz.wikipedia.org>
11. <https://ares.uz>
12. <https://ziyocom.uz>

QORAQALPOQ VA O‘ZBEK TILLARINING SINTAKTIK MODELLARINI TEGGLASH MASALASI

Munisa Xudoyberganova

Mirzo Ulug‘bek nomidagi

O‘zbekiston Milliy universiteti 2-bosqich magistri

E-mail: munisaxudayberganova13@gmail.com

Annotatsiya: Ushbu maqolada Universal Dependencies (UD) tizimi va uning o‘zbek hamda qoraqalpoq tillaridagi qo‘llanilishi yoritiladi. Teglash jarayoni va uning tabiiy tilni qayta ishlash (NLP)dagi ahamiyati, xususan tokenlash, lemmatizatsiya va stemmizatsiya kabi bosqichlar bilan bog‘liq jihatlar ko‘rib chiqiladi. UD tizimining so‘z turkumlari, morfologik va sintaktik xususiyatlarni belgilashda qanday ishlashi, ayniqsa o‘zbek tilidagi so‘z birikmalarini teglashdagi yondashuvlar tahlil qilinadi.

Kalit so‘zlar: Teglash, sintaktik birlik, so‘z birikmalari, Universal Dependencies (UD)

Abstract: This article explores the Universal Dependencies (UD) system and its application in the Uzbek and Karakalpak languages. It analyzes the tagging process and its significance in Natural Language Processing (NLP), especially in stages such as tokenization, lemmatization, and stemming. The paper describes how UD is used to define parts of speech, morphological, and syntactic features, particularly in the annotation of word combinations in Uzbek.

Keywords: Tagging, syntactic unit, word combinations, Universal Dependencies (UD)

Аннотация: В данной статье рассматривается система Universal Dependencies (UD) и её применение в узбекском и каракалпакском языках. Анализируется процесс аннотирования (теггирования) и его значение в области обработки естественного языка (NLP), особенно в таких этапах, как токенизация, лемматизация и стемминг. Описано, как система UD используется для определения частей речи, морфологических и синтаксических характеристик, в том числе при аннотировании словосочетаний в узбекском языке.

Teglash inglizcha so‘z bo‘lib, tag –“belgi, yorliq” ma’no iudasini beradi. Bu jarayon matn, tasvir yoki boshqa tur ma’lumotlarni identifikatsiya qilish uchun qo‘llaniladi. Teglashning asosiy maqsadi axborotni turkumlash va aniq tasniflashdir. Ma’lumotlar tez va samarali qidirish uchun tizimlanadi, kontent mazmuni qayta ishlanadi, veb-sahifalarning qidiruv tizimlarida yaxshi indekslanishiga yordam berish, sun’iy intellekt yoki kompyuter tizimlariga tushunarli bo‘lishini ta’minlashdan iboratdir.

Teglash texnologiyalar, ilmiy tadqiqotlar va internetda ma’lumotni tashkil qilishning ajralmas qismi hisoblanadi. U ma’lumotlarni tezkor topish, to‘g‘ri tahlil qilish va avtomatlashtirishda samarali vosita hisoblanadi.

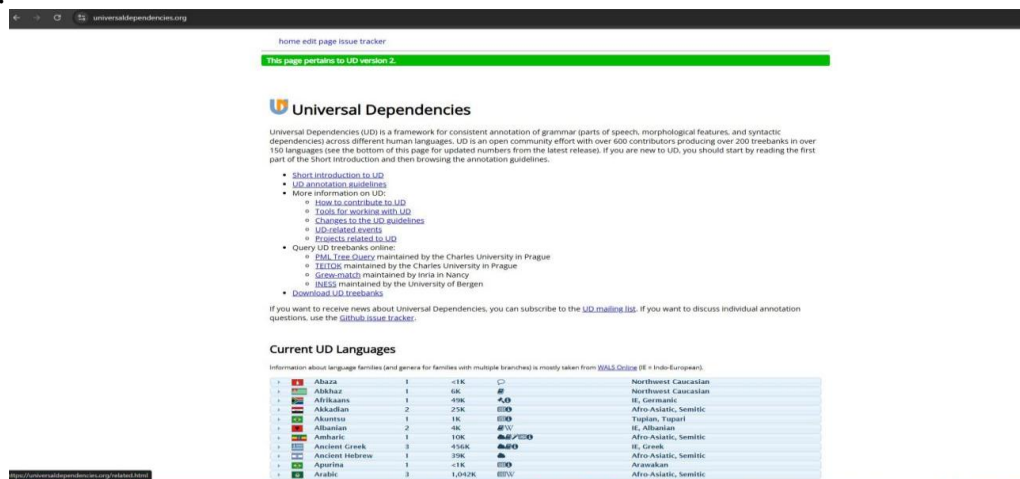
Matnlarni me‘yorlashtirish jarayoni morfologik tahlil bilan bog‘liqdir. Bu esa quyidagi bosqichlarda amalga oshiriladi:

- Tokenlash – matndagi so‘z shakllarini aniqlash;
- Lemmatizatsiya – so‘zlarning lug‘atdagi ifodasini aniqlash;
- Stemimizatsiya – derivative so‘zlarning o‘zagini aniqlash.

So‘ngi yutuqlar tabiiy tilni qayta ishlash (NLP) sohasida har qanday tilga, ayniqsa resurslari kam bo‘lgan tillarga qo‘llanilishi mumkinligini ko‘rsatdi. Universal Dependencies¹³⁷ shunday kross-lingvistik loyihalardan biridir, uning maqsadi annotatsiyalangan lingvistik

¹³⁷ <https://universaldependencies.org/>

ma'lumotlarni to'plashdir va bu qoidalar, ideal holda, har qanday tilda, universal tarzda qo'llanilishi va mos kelishi kerak. NLP va lingvistika sohalaridagi kross-lingvistik tadqiqotlar va qo'llanmalarda, bir xil annotatsiya qoidalari bilan annotatsiyalangan lingvistik ma'lumotlarni ishlatish juda muhimdir¹³⁸. Ammo, amalda, hozirgi UD daraxt bazalarida annotatsiya qoidalari bo'yicha bir qator kelishmovchiliklar mavjud, hatto bir xil tilda ham. Ushbu annotatsiya uslubidagi nomuvofiqlik bir nechta daraxt bazalarining bir vaqtning o'zida ishlatilishini qiyinlashtiradi, chunki bu, masalan, nazoratli vazifalar va transfer o'rganishda chalkashliklarga olib kelishi mumkin. Universal Dependencies (UD) – bu turli tillarda grammatika (so'z turkumlari, morfologik xususiyatlar va sintaktik bog'lanishlar)ni izchil annotatsiya qilish uchun mo'ljallangan bir tizimdir.



3-rasm. Universal dependencies platformasi.

UD maqsadi — tillar orasida bir xil annotatsiya qoidalarini qo'llash va buni tabiiy tilni qayta ishlash (NLP) va lingvistik tadqiqotlar uchun foydali qilishdir. UDning asosiy tamoyili shundaki, u tilga xos bo'lgan morfologik va sintaktik xususiyatlarni tushunishda yordam beradi va shu bilan birga ko'plab tillar uchun umumiy qo'llaniladigan qoidalarini yaratadi¹³⁹. Misol uchun, UD yordamida bir fe'lning "bosh fe'l" yoki "yordamchi fe'l" ekanligini, yoki bir so'zning nima ma'no yoki rolni bajarishini belgilash mumkin. Bu ma'lumotlar yordamida, masalan, kompyuterlar turli tillarda matnni tahlil qilish yoki tarjima qilishda yordam berishi mumkin. UDning afzalligi shundaki, u bir nechta tillarni taqqoslash va ularning grammatik tuzilmalarini bir-biriga moslashtirishga imkon beradi. Bu nafaqat tadqiqotlar uchun, masalan, tilni avtomatik ravishda tahlil qilish yoki mashina tarjimasini sohasida ham foydalidir. UD tizimi quyidagi so'z turkumlarini belgilashda umumiy kategoriyalarni taklif qiladi¹⁴⁰:

Traditional POS	UPOS	Category
noun	NOUN	common noun
	PROPN	proper noun
verb	VERB	main verb
	AUX	auxiliary verb or other tense, aspect, or mood particle
adjective	ADJ	adjective
	DET	determiner (including article)
	NUM	numeral (cardinal)
adverb	ADV	adverb
pronoun	PRON	pronoun
preposition	ADP	adposition (preposition/postposition)
conjunction	CCONJ	coordinating conjunction
	SCONJ	subordinating conjunction
interjection	INTJ	interjection
—	PART	particle (special single word markers in some languages)
—	X	other (e.g., words in foreign language expressions)
—	SYM	non-punctuation symbol (e.g., a hash (#) or emoji)
—	PUNCT	punctuation

4-rasm. UD tizimidagi teglash kategoriyasi.

¹³⁸ Ch, Taguchi. Consistent Grammatical Annotation of Turkic Languages for more Universal Dependencies.

¹³⁹ https://www.researchgate.net/publication/350477280_Universal_Dependencies.

¹⁴⁰ https://www.researchgate.net/publication/350477280_Universal_Dependencies.

Har bir tilning o'ziga xos xususiyati bor. Ushbu kategoriyalar dunyodagi tillarda keng tarqalgan. Biz barcha tillar ushbu kategoriyalarning barchasidan foydalanishi kerak, degan ikrdan uzoqmiz, ammo tildagi har bir so'zni ulardan biriga tayinlash mumkin deb hisoblaymiz. Ba'zi so'z sinfini ajratishlar UDda ayniqsa muhimdir: Masalan, otlar va fe'llar o'rtasidagi ajratish, tarkibiy qismning nominal yoki klauzal ekanligini aniqlashda muhim rol o'ynaydi.

Quyidagi rasm Universal Dependencies (UD) tizimiga asoslangan va O'zbek tili bo'yicha turli grammatika, so'z turlari va ularning sintaktik xususiyatlari haqida ma'lumot beradi.

UD for Uzbek



Tokenization and Word Segmentation

- In general, words are delimited by whitespace characters.
- According to typographical rules, many punctuation marks are attached to a neighboring word. These are normally tokenized as separate tokens (words), with the following exception:
 - The period that marks an abbreviation is part of the abbreviation token: *mln.* "million", *A. Navoiy* "A. Navoi"

Morphology

- Uzbek has a rich inflectional and derivational morphology.
- Question suffixes are written adjacent to the word in Uzbek. *-mi*

Tags

- Uzbek uses all 17 UPOS categories:
 1. PRON: Men, sen, u.
 2. PUNCT: -, ., !, ?
 3. VERB: o'qidi, kelmoqchi, ishla.
 4. NOUN: oila, do'st, quyosh.
 5. AUX: kerak, mumkin, edi, emas.
 6. ADP: bilan, uchun, haqida.
 7. CCONJ: va, yoki, hamda.
 8. SCONJ: agar, chunki, ya'ni.
 9. ADV: atayin, juda, tez.
 10. PROPN: Toshkent, Navoiy, Bobur.
 11. DET: bu, ayrim, barcha.
 12. ADJ: qizil, tinch, aqlli.
 13. NUM: ikki, yuz, ming.
 14. INTJ: hoy, salom, ofarin.
 15. PART: -mi, -chi, -a/-ya
 16. SYM: \$, +, =
 17. X

5-rasm. UD tizimidagi O'zbek tili uchun teglash kategoriyasi.

Quyida matndagi asosiy tushunchalarni tahlil qilamiz:

1. **Tokenlash va so'zni segmentlash:** O'zbek tilida so'zlar asosan bo'sh joy yordamida ajratiladi. Ba'zi belgilarga (masalan, nuqta yoki vergul) yaqin joylashgan so'zlar odatda alohida token (so'z) sifatida ajratiladi, ammo qisqartmalar (masalan, "mln." yoki "A. Navoiy") bitta token sifatida qabul qilinadi.

2. **Morfologiya:** O'zbek tili boy morfologiyaga ega, ya'ni so'zlar ko'p holatlarda o'zgaradi. Masalan, savol qo'shimchasi (misol: "-mi") so'zga qo'shib yoziladi.

3. **So'z turlari (UPOS):** O'zbek tilida so'zlar 17 turga bo'linadi (ADJ, ADP, ADV, AUX, CCONJ, DET, INTJ, NOUN, NUM, PART, PRON, PROPN, PUNCT, SCONJ, SYM, VERB, X). Masalan, **PRON** (olmosh), **VERB** (fe'l), **NOUN** (ot), **ADJ** (sifat), **NUM** (son), va boshqalar). **PRON** turidagi so'zlar: "Men", "Sen", "U"; **VERB** turidagi so'zlar: "o'qidi", "kelmoqchi", "ishla"; **NOUN** turidagi so'zlar: "oilal", "do'st", "quyosh".

4. **Nominal xususiyatlar:** O'zbek tilida 6 ta holat (Nom, Acc, Gen, Dat, Loc, Abl) mavjud. Son (Singular va Plural) faqat ko'plik qo'shimchasi mavjud bo'lsa, ko'plik sifatida belgilangan.

5. **Fe'l xususiyatlari:** Fe'llar **finita** (kelishilgan) yoki **non-finita** (kelishilmagan) bo'ladi. Fe'llar **Aspect** (holat), **Mood** (kayfiyat), **Tense** (zamon) va boshqa xususiyatlar bilan aniqlanadi.

6. **Leksik xususiyatlar:** Olmoshlar va belgilovchilar turli turlarga bo'linadi, masalan: shaxsiy (Prs), demonstrativ (Dem), noaniq (Ind) va boshqalar.

7. **Sintaksis (Syntax):** O'zbek tilida asosiy so'z tartibi **Subyekt-Obyekt-Fe'l** bo'ladi. Nominal subyektlar (nsubj) va obyektlar (obj) holatlarni bildiradi, qo'shimchalar (adjuncts) esa qo'shimcha ma'lumot beradi.

8. **Munosabatlar:** O'zbek tilida turli sintaktik munosabatlar mavjud, masalan:

- **advcl:cond** - shartli adverbial klauzalar (masalan, "Agar u kelmasa, biz ketamiz").
- **nmod:poss** - egalik modifikatorlari (masalan, "bobomning uyi").
- **compound:lvc** - engil fe'l konstruksiyalari (masalan, "yo'lga chiqmoq").

Bu ma'lumotlar o'zbek tilidagi so'zlarni to'g'ri tahlil qilish va belgilashga yordam beradi, shuningdek, tilni yaxshiroq tushunish va tahlil qilish imkonini yaratadi.

Universal Dependencies (UD) tizimi so'z birikmalarini teglashda ham ishlatiladi. So'z birikmalarini to'g'ri teglash uchun, so'zlar o'rtasidagi sintaktik munosabatlarni aniqlash kerak. UD tizimida, so'z birikmalari (compound so'zlar, o'zaro bog'langan nomlar va boshqa sintaktik tuzilmalar) uchun maxsus teglar mavjud. O'zbek tilida so'z birikmalarini teglashda quyidagi usullarni qo'llash mumkin:

- **Compound so'zlar:** Agar birikmada ikki yoki undan ortiq so'zlar birgalikda bitta ma'no yaratgan bo'lsa (masalan, "Toshkent shahar"), bunday birikmalar **compound** tegini oladi. Bu holatda, birikmadagi har bir so'z uchun alohida token (so'z) qo'llanadi va ular **compound** munosabati bilan bog'lanadi. **Masalan**, "Toshkent shahar".
 1. Toshkent — **PROPN** (maxsus nom)
 2. shahar — **NOUN** (ot)
 3. Munosabat: **compound** (so'zlar birikmasi)
- **Nominal modifikatorlar (nmod):** Nominal modifikatorlar (masalan, "bobomning uyi") o'zgaras bo'lishi mumkin, shuning uchun ular **nmod:poss** (egallik) munosabati bilan taglanadi. **Masalan**, "bobomning uyi".
 1. bobomning — **NOUN** (ot)
 2. uysi — **NOUN** (ot)
 3. Munosabat: **poss** (egallik)
- **Adverbial modifikatorlar (advmod):** Agar birikmada adverb (ravish) va boshqa so'z biriksa, **advmod** tegi ishlatiladi. Masalan, "tezda borish".
 4. tezda — **ADV** (ravish)
 5. borish — **VERB** (fe'l)
 6. Munosabat: **advmod** (ravishning modifikatori)
- **Adjectival Modifikatorlar (amod):** Agar sifat (adjective) va ot biriksa, ular **amod** (adjectival modifier) munosabati bilan bog'lanadi. **Masalan**, "qizil gul".
 7. qizil — **ADJ** (sifat)
 8. gul — **NOUN** (ot)
 9. Munosabat: **amod** (sifat modifikatori)
- **Compound** so'zlar bir-biriga bog'langan va bitta ma'no hosil qilgan so'zlardan iborat bo'ladi (masalan, "yangi yil").
- **nmod** tegini ishlatish, asosan, otlar o'rtasidagi bog'lanishni ko'rsatadi, ayniqsa egallik va modifikatsiyalarni ifodalashda.
- **advmod** va **amod** teglarini esa tegishli ravish va sifatlarni modifikator sifatida ishlatish mumkin.

Xulosa qilib aytganda, so'z birikmalarini to'g'ri teglash uchun har bir so'zga uning grammatik funksiyasiga mos teg beriladi va o'rtadagi sintaktik munosabatlar aniqlanadi. UD tizimi bunday sintaktik tuzilmalarni aniq belgilash va bir-biriga bog'lash imkonini beradi. Qoraqalpoq va o'zbek tillarining sintaktik modellarini teglashda UD tizimidagi teglardan foydalanish maqsadga muvofiqdir. O'zbek va qoraqalpoq tillari sintaktik modellarini teglashda **Universal Dependencies** tizimi ikkala tilda ham qo'llaniladi. Ikkala tilning sintaktik tuzilmalari o'zaro

o'xshash bo'lsa-da, ayrim morfologik farqlar mavjud. Biroq, umumiy sintaktik modellash va teglash usullari o'zaro muvofiqlashgan bo'lib, bu tillarni samarali tahlil qilishda va taqqoslash mumkin bo'ladi

Foydalanilgan adabiyotlar:

1. Ch, Taguchi. Consistent Grammatical Annotation of Turkic Languages for more Universal Dependencies.
2. Abduraxmonova N. Mashina tarjimasining lingvistik asoslari. – T., 2012.
3. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
4. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
5. Alkim.E, Çebi. Y. Machine Translation Infrastructure for Turkic Languages (MT-Turk).– Turkiya-2019.
6. Benita Pungwe Mpala. THE USE OF ADVANCED TECHNOLOGY TO FIND SOLUTIONS FOR HUMANITARIAN ISSUES.–San Fransisko, 2020. P–20.
7. Brown,P.F., Cocke, J., Della Pietra, S.A., Mercer, R. L. The Mathematics of Statistical Machine Translation:Parameter Estimation. Computational linguistics, 19(2),–P.263-311.

TEZAURUS VA TEZAURUS LUG'ATLARNI YARATISHGA DOIR

Obidjonova Odina Obidjon qizi

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti
Jurnalistika va o'zbek filologiyasi fakulteti 2- kurs talabasiE-mail: odinaobidjonova03@gmail.com

Annotatsiya: Mazkur maqolada tezaurus haqida umumiy ma'lumotlar va uning turlari, foydalanish maqsadlari haqida turli ilmiy manbalar va tadqiqotlardan unumli foydalanilgan holda ma'lumotlar keltirilgan. Shuningdek maqolada sifat turkumiga oid "xursand" so'zining izohlari, uning so'zlarning ma'no munosabatiga ko'ra turlari doirasida qanday graduonimik qatorda turishi, so'zlar bilan birikma hosil qilishi haqida fikr yuritilgan. Shuningdek, ba'zi izoh va so'zlar gaplar misolida tushuntirib berilgan.

Kalit so'zlar: tezaurus lug'atlar, tezaurus turlari, umumiy tezaurus, mavzuga xos tezaurus, iyerarxik tezaurus, assotsiativ tezaurus, onlayn tezaurus, sinonim, antonim, graduonimik qator.

Абстрактный: В данной статье представлены общие сведения о тезаурусе, его разновидностях и целях использования на основе различных научных источников и исследований. Кроме того, в статье рассмотрены толкования прилагательного "xursand" (радостный), его место в градуонимическом ряду в зависимости от семантических отношений между словами, а также его сочетаемость с другими словами. Некоторые толкования и слова были разъяснены с использованием примеров в предложениях.

Ключевые слова: тезаурусные словари, виды тезаурусов, общий тезаурус, тематический тезаурус, иерархический тезаурус, ассоциативный тезаурус, онлайн-тезаурус, синоним, антоним, градуонимический ряд.

Annotation: This article provides general information about the thesaurus, its types, and purposes of use, utilizing various scientific sources and research effectively. Additionally, the article discusses the definitions of the adjective "xursand" (cheerful), its position within a gradonymic scale based on semantic relations between words, and its ability to form collocations with other words. Some definitions and words are also explained through example sentences.

Keywords: thesaurus dictionaries, types of thesauri, general thesaurus, subject-specific thesaurus, hierarchical thesaurus, associative thesaurus, online thesaurus, synonym, antonym, gradonymic scale.

XXI asr kompyuter texnologiyalari asri, butun dunyoda raqamlashtirish jarayoni jadallik bilan rivojlanib borayotgan davr. Bu davrda kompyuter lingvistikasining bir qancha yo'nalishlari bilan birgalikda kompyuter leksikografiyasi, tezaurus, ontologik lug'atlar ham rivojlanib bormoqda. Tezaurus lug'atlari haqida gapiradigan bo'lsak, u tilshunoslikdagi leksik birliklar o'rtasidagi omonimik, sinonimik, paronimik, antonimik, gipo-giperonimik (tur-jins), partonimik va boshqa munosabatlarni ifodalaydigan maxsus lug'at turidir. Tezaurus ma'lum bir tildagi barcha so'zlar vositasida semantik aloqalarni o'zida aks ettiradigan so'zlar jamlanmasi hisoblanadi. Tezauruslar ilmiy, sohaviy bo'lishi ham mumkin. Masalan, fan, iqtisodiyot, tibbiyot sohasida yaratilgan tezaurus lug'atlar yoki huquq, filologiya yo'nalishlarida yaratilgan tezaurus lug'atlar (dialektal lug'atlar va h) bo'lishi mumkin.

Tezaurus muayyan tilda barcha so'zlarni qamrab oladigan, ularning matnda qo'llanish holatlarini to'la-to'kis aks ettiradigan lug'at. Muayyan yozuv yodgorliklaridan leksik birliklarni yoppasiga terib olishga asoslangan (masalan, yunon, lotin tillarida tuzilgan) lug'atlar shunday lug'atlardan hisoblanadi. So'z tanlash tamoyillariga ko'ra yozuvchilar yoki ularning ijodiga mansub biror asar tili bo'yicha tuzilgan lug'atlar ham tezaurus hisoblanadi. U leksik-semantik,

korporativ kommunikatsiya (bir fan yoki kasb orqali o'zaro bog'liq bo'lgan shaxslarning muloqotda bir-birlarini tushinishlari) uchun xizmat qiladi [B.Suyunov].

Tezaurus ma'lum bir so'zning qidiruv tizimi orqali omonim, sinonim, paronim, antonimlarini topishda yordam beradi, nafaqat semantik aloqalarni balki o'sha so'zlarning qanday so'zlar bilan birikib kela olishi, ibora holatini, maqollar, hikmatli so'zlar ichida kelishini ko'rsatib beradi. Bu jihat tadqiqotchilar uchun foydali va samarali.

Kompyuter lingvistikasining tarkibiy qismi bo'lgan tezauruslar leksemalarning o'z va ko'chma ma'nolari, kommunikativ xususiyatlari, emotsional-ekspressiv vazifalari haqida keng ma'lumot berishi bilan qimmatlidir [Kutuzov].

Tezaurus yunoncha so'zdan olingan bo'lib, "xazina, boylik" degan ma'nolarni anglatadi. Birinchi zamonaviy ingliz tezaurusi Piter Mark Roger tomonidan 1805-yilda yaratilgan. Mazkur tezaurus 1852-yilda Angliyada nashr etilgan, taxminan 15 mingta tushunchani o'z ichiga oladi va shu vaqtdan beri foydalanib kelinadi. Bunday lug'atlarning asosiy maqsadi matn yozishda tegishli so'zlarni tanlashga yordam berish, mavzu doirasida atama yoki termin o'rtasidagi munosabatlarni tavsiflovchi ma'lumotlarni qidirib topishdan iborat. Bunday lug'atlar ma'lum bir soha mutaxassislari tomonidan yaratilgan va mavzu doirasidagi ma'lumotni qidirib topish uchun mo'ljallangan. 1870-yillarda tezauruslardan ma'lumot izlash ishlarida faol foydalanila boshlandi. Shunday lug'atlardan yana biri Vikilug'at deb ataladi. U ko'p tilli bepul yangilanadigan lug'at va tezaurus bo'lib, "Vikimediya" fondi loyihasi doirasida yaratilgan va 2004-yildan hozirgi kungacha ishlatilib kelinadi. Unda 250 mingta tushuncha va 67 mingta semantik munosabatlar o'z aksini topgan.

Tezaurus yaratish borasida xorijda muhim tadqiqotlar olib borilgan. Ingliz tilidagi lug'at-tezaurus, V.Dal tomonidan yaratilgan izohli lug'at-tezaurus, T.F.Yefremovning rus tilidagi yangi so'zlarning lug'at tezaurusi, S.I.Ojegov, N.Y.Shvedova tomonidan yaratilgan rus tilining izohli lug'ati-tezaurusi, Fasmer Maksning rus tilining etimologik lug'ati-tezaurusi, yuridik terminlar tezaurusi yaratilgan.[3]

Tezaurus bir nechta maqsadlarda foydalaniladi:

1.Lug'at boyligini kengaytirish: Ular foydalanuvchilarga yozish va nutqini yaxshilash uchun muqobil so'z va iboralarni topishga yordam beradi, bu esa muloqotda aniqlik, rang-baranglik va ravshanlikni ta'minlaydi.

2.Sinonimlarni qidirish: Tezaurus foydalanuvchilarga so'z tanlashda qiyinchiliklarga duch kelganda yoki yozishda takrorlanishdan qochishda ya'ni sinonimlarni topishda yordam beradi.

3.So'rovlarni kengaytirish: Axborot qidirish tizimlarida tezaurus so'rov shartlarini sinonimlar va tegishli tushunchalarni o'z ichiga olgan holda kengaytirish orqali qidiruv natijalarini yaxshilashi mumkin va shu bilan tegishli ma'lumotni topish ehtimolini oshiradi.

4.Matn tasnifi: Tezaurus lug'atlari matnni tahlil qilish va tasniflash vazifalarini bajaradi, matn ma'lumotlarini turkumlash va tartibga solishda yordam beradi.

5.Kontsepsiya xaritasi: Tezaurus so'zlar va tushunchalar o'rtasidagi kontseptual aloqalarni xaritalashni osonlashtiradi, bilimlarni tashkil qilish va kontsepsiyani vizualizatsiya qilishga yordam beradi.

6.Til o'rganish: Tezaurus til o'rganuvchilarga muqobil so'z va iboralarni taqdim etish orqali yordam beradi, ularning so'z boyligini kengaytirishga va nozik ma'nolarni tushunishga yordam beradi.

7.Shuni ta'kidlash kerakki, tezauruslar qimmatli vositalar bo'lishi mumkin bo'lsa-da, ulardan foydalanish puxta o'ylash va kontekstni yaxshi tushunishni talab qiladi. Eng mos sinonim yoki tegishli atamani tanlash ma'no va foydalanishdagi nozik farqlarni tushunishni talab qiladi.

So'zlarni to'g'ri tanlashni ta'minlash uchun lug'atlar va boshqa tillarga murojaat qilish tavsiya etiladi.

8.Eng muhim vazifasi shundan iboratki, tezaurus so'zlarning sinonimlari va antonimlarini taqdim etadigan, foydalanuvchilarga muqobil iboralarni taklif qiladigan va so'z boyligini kengaytirish, so'z tanlash, ma'lumot qidirish va til o'rganishga yordam beradigan manbadir.

Tezauruslarning bir nechta turlari mavjud bo'lib, ularning har biri yaratilish sababiga, foydalanish auditoriyasiga qarab turli maqsadlarda foydalaniladi. Jumladan:

1. Umumiy tezaurus: tezaurusning bu turi keng lug'atni qamrab oladi va umumiy tildan foydalanish uchun mo'ljallangan. U bir nechta domenlardagi so'zlarning sinonimlari va antonimlarini taqdim etadi va kundalik tilga murojaat qilish va lug'atni kengaytirish uchun mo'ljallangan.

2. Mavzuga xos tezaurus: Ushbu turkumdagi tezauruslar tibbiyot, huquq yoki texnologiya kabi ma'lum bir mavzu yoki sohaga qaratilgan. Ular qiziqish sohasiga xos ixtisoslashgan terminologiya, sinonimlar va tegishli atamalarni taqdim etadi. Ushbu tezauruslar sohaga oid lug'at va aniq atamalarni taklif qilish orqali o'z sohalarida mutaxassislar, tadqiqotchilar va talabalarga yordam berish uchun mo'ljallangan.

3. Iyerarxik tezaurus: tezaurusning bu turi so'zlarni iyerarxik tuzilishda tartibga soladi, bu yerda kengroq terminlar (giperonimlar) va torroq terminlar (giponimlar) olinadi. Bundan tashqari, meronim (qism orqali butun tushuniladi), xolonim (butun qismi orqali qism tushuniladi), partonim (qism) kabi iyerarxik tushunchalar atamalar orasidagi iyerarxik munosabatlarni ko'rsatib, foydalanuvchilarga tushunchalar bo'ylab harakatlanish va tilning iyerarxik tuzilishini tushunish imkonini beradi.

4. Assotsiativ tezaurus: Assotsiativ tezaurus qattiq iyerarxiyadan ko'ra so'zlar orasidagi assotsiativ munosabatlarga urg'u beradi. Ular so'zlarni semantik assotsiatsiyalar asosida birlashtiradi, masalan, odatda birga topiladigan yoki o'xshash kontekstlarni o'rtoqlashadigan so'zlar. Ushbu tezauruslar foydalanuvchilarga iyerarxik jihatdan bog'liq bo'lmasa ham, kontseptual aloqaga ega bo'lgan so'zlarni o'rganishga yordam beradi.

5. Onlayn tezaurus: Raqamli texnologiyaning paydo bo'lishi bilan tezauruslar endi onlayn formatlarda mavjud. Onlayn tezauruslar so'zlar, sinonimlar va tegishli atamalarning keng ma'lumotlar bazalariga dinamik kirishni ta'minlaydi. Ular ko'pincha qidiruv imkoniyatlari, xatcho'plar va o'zaro havolalar kabi qo'shimcha funksiyalarni o'z ichiga oladi, bu ularni til ma'lumoti uchun qulay va foydalanuvchilarga qulay manbalar qiladi.

Vaitova ta'kidlaganidek, tezaurus termini, avvalambor, tilning lug'at fondini maksimal darajada to'liqlik bilan ifodalaydigan xazina sifatida talqin etiladi. Matnlarning adekvat talqini va ma'lumotlari uning sistem ma'nolarini o'zida mujassam etgan va soha vakillari tomonidan e'tirof etilgan tezaurusda aniqlanadi. Tezaurus atamasini L.V.Shcherba esa shunday izohlaydi: "Tezaurus deyishganda, hozirgi paytda ular ko'pincha "Thesaurus linguae latinae" ni nazarda tutadilar, ular beshta nemis akademiasining korxonasi bo'lib, 1900-yilda boshlangan. Xarakterli xususiyati: ma'lum bir tilda ro'y bergan barcha hal qiluvchi so'zlarni o'z ichiga oladi va har bir so'z ostida ushbu tilda mavjud bo'lgan matnlardan misollar keltiriladi" [J.I.B.IIqep6a, 1940:10]. Vaitova va L.V.Shcherba aytib o'tganidek, tezaurus tildagi barcha so'zlarning lug'atini qamrab olgan, o'sha tilning lug'at fondini to'liqlik bilan ifodalaydigan xazinadir, tezaurusda lug'atda mavjud so'zlarga turli matnlardan misollar keltiriladi. Masalan, O'zbek tilining elektron korpusida mavjud tezaurus lug'atida so'zlarni so'z bo'yicha yoki guruh bo'yicha qidirishimiz mumkin. Ushbu korpusdagi tezaurusda so'zlarni so'z bo'yicha qidirganimizda so'zlarning o'zbekcha, ruscha, inglizcha varianti, so'zning izohlari va ushbu so'z ishtirok etgan matnlardan namunalar keltiriladi, so'zlarning sinonim, omonim, paronim, antonimlari, giponim, giperonimlari, meronim-xolonimlari chiqarib beriladi. Shuningdek, tezaurus lug'atida o'sha so'z qatnashgan maqollar, iboralar va u bilan birika oladigan so'zlar ham birikma holda beriladi. O'zbek tilining elektron korpusi tezaurus lug'atidan anchagina so'zlar joy olgan, ammo ko'p so'zlar hali ham lug'atga kiritilmagan. Bu anchagina ko'p vaqt va mehnatni talab qiladi. Lekin bu borada yurtimizda jadal ishlar olib borilmoqda.

O'zbek tilshunosligida ham kompyuter leksikografiyasi yuzasidan tadqiqotlar olib borilmoqda. O'zbek tilidagi har bir so'zning tezaurus ko'rinishini ochib berish qiyin vazifa bo'lganligi sababli o'zbek tilshunosligida fikrning mohiyatini ifodalaydigan, gapning markazini tashkil etuvchi fe'l so'z turkumi yuzasidan so'zlar jamlanib tadqiq etilgan. O'zbek tilshunosligida

tezauruslarning yaratilishida o'zbek tili xususiyatlarini hisobga olish lozim bo'ladi. Har bir til birligiga oid tushunchalar tezaurusini berishda ularning nutqdagi ahamiyatlilik darajasidan kelib chiqib yondashiladi. Turkiy tillarda, jumladan, o'zbek tilida fe'llar fikr ifodalashda muhim o'rin tutadi. Shu bois o'zbek tilidagi fe'llar tezaurusini boshqa tillarga qiyoslagan holda yaratish tilning ijtimoiy funksiyasini oshirishda, axborot uslubini shakllantirishda va takomillashtirishda muhim o'rin tutadi.[3]

Ushbu maqolani yozishdan maqsad tezaurus lug'ati va uning maqsadi, turlari, bu borada qanday ishlar amalga oshirilganligi haqida ma'lumot berish va sifat turkumidagi "xursand" so'zining tezaurus lug'atini yaratishdir.

Xursand

Радостный

Cheerful

Izoh: Hayotdan juda mamnun, kayfiyati juda yaxshi

Namuna: Ozod yoshlarning ruhlari tetik va o'zlari xursand edilar. N. Fozilov, Diydor

Oyim azbaroyi xursand bo'lganidan: — Barakalla, o'g'lim! — deb yelkamga qoqib qo'ydi.

H. Nazir, Yonar daryo

Izoh: Biror kimsadan yoki narsadan qanoat hosil qilgan; mamnun, rozi, minnatdor

Namuna: Shu yo'sin Nodira Mohlaroyimdan juda xursand bo'lgani holda uyiga qaytdi. A.

Qodiriy, Mehrobdan chayon

Elmurod uni [Davlatyorni] shunday sharoitda uchratganidan xursand bo'ldi. P. Tursun,

O'qituvchi

Izoh: Xursand (erkaklar va xotin-qizlar ismi)

Namuna: Salim xonaga endi kirgan ham ediki, Xursand aka kelib qoldi.

Sinonimik qatori: xursand, xushvaqt, baxtiyor, shod, xurram, xushnud, masrur, sarafroz

Bosh so'z: Xursand

Kontekstual sinonimlari: Mamnun, rozi, minnatdor, baxtli

Namuna: U o'qishga kirganidan xursand edi. U o'qishga kirganidan mamnun edi.

Ota o'g'lining bu ishlaridan xursand edi. Ota o'g'lining bu ishlaridan rozi edi.

Shoikrom ukasining bu yordamidan juda ham xursand edi. Shoikrom ukasining bu yordamidan juda ham minnatdor edi.

Dilafruz yigitning bu ehtirosli so'zlaridan ichiga sig'may, baxtli edi. Dilafruz yigitning bu ehtirosli so'zlaridan ichiga sig'may, xursand edi.

"Xursand" so'ziga teng sinonim iboralar: Og'zi qulog'iga yetmoq, boshi ko'kka yetmoq, terisiga sig'may ketmoq, qo'yi mingga yetmoq, og'zining tanobi qochmoq, dimog'i chog' bo'lmoq, ko'ngli ko'tarilmoq, ko'ngli yorishmoq

Antonimlar: xafa, g'amli, qayg'uli, hasratli, tushkun, g'amgin, ma'yus, dardli, xomush, o'ksik

"Xafa" so'ziga teng sinonim iboralar: ko'ngli o'ksigan, ko'ngli og'rigan, ko'ngli yara, yuragi siqilgan, ko'ngli siniq, ko'ngli o'ksigan, yuragiga og'ir botgan, dili siyoh bo'lgan, dili og'rigan, diliga ozor yetgan, g'amga botgan, ruhi tushgan, kayfiyati tushgan, ko'ngli qolgan, yuragi ezilgan, kayfi buzuq, ta'bi xira, ta'bi tirriq, avzoyi buzuq, qovog'idan qor yog'gan

Omonim: yo'q

Paronim: yo'q

Iboralar: yo'q

Birikmalar: tanishganidan xursand, ota-onasi uchun xursand, juda xursand bo'lmoq, muvaffaqiyatidan xursand, hayotdan xursand, xushxabardan xursand, tug'ilganiga xursand, yashayotganidan xursand, sog' qolganiga xursand, o'qituvchilaridan xursand

Yasalishiga ko'ra: sodda sifat

Turiga ko'ra: asliy sifat (ma'noviy turiga ko'ra: xususiyat sifati)

Birikma yasalish qolipi:

xursand + ot so'z turkumi = bitishuv munosabatli birikma

Masalan: xursand bola, xursand qiz

Leksema (-dan) + xursand = boshqaruv munosabatli birikma

Masalan: hayotdan xursand, tanishganidan xursand

Soʻz yasalihi qolipi:

xursand + boʻlmoq, qilmoq, etmoq = qoʻshma feʼl

Xursand + (-)chilik, -lik = ot soʻz turkumi

Xulosa. Tezaurus bu soʻz boyligini oshirish, semantik tahlil qilish va maʼlumotlarni qidirishda yordam beradigan maxsus lugʻatdir va u muayyan bir tilda yaratilgan turli sohadagi soʻzlar yigʻindisi hisoblanadi. Tezaurus yuqorida aytib oʻtganimizdek, turli maqsadlarda foydalaniladi. Masalan, til oʻrganish, sinonimlarni qidirish, lugʻat boyligini kengaytirish kabi. Ushbu maqolada, asosan, "xursand" soʻzining tezaurus lugʻatini yaratishga keng eʼtibor qaratildi. Biz bilamizki, "xursand" soʻzi sifat soʻz turkumiga mansub va u sinonimlar qatorida dominant soʻz hisoblanadi. "Xursand" soʻzi hissiy boʻyogʻiga koʻra ijobiy boʻyoqdor soʻzdir. Maqolada ushbu soʻzning tezaurus lugʻatiga xos jihatlari ochib berildi.

Foydalanilgan adabiyotlar roʻyxati:

1. N. Abdurakhmonova, I. Alisher and G. Toirova, "Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing," 2022 7th International Conference on Computer Science and Engineering (UBMK), Diyarbakir, Turkey, 2022, pp. 73-75, doi: 10.1109/UBMK55850.2022.9919521.
2. Suyunov B.T. Tezaurus va kompyuter texnologiyalariga doir // "KOMPYUTER LINGVISTIKASI: MUAMMOLAR, YECHIM, ISTIQBOLLAR" Respublika I ilmiy-texnikaviy konferensiya. Vol. 1 №. 01 (2021)
3. Poʻlatova G.Gʻ. Oʻzbek tili korpusi uchun tezaurus lugʻatlar yaratish. zaruriyati <https://t.me/openscholar> Multidisciplinary Scientific Journal January, 2024. ISSN: 2181-4147 VOLUME 2 | ISSUE 1 | 2024
4. Oʻzbek tilining izohli lugʻati.- 2025 // <https://sinonim.uz/words/xursand>
5. Saporova, M. F. Lingvistik tezaurus yaratish tajribasidan // Soʻz sanʼati: xalqaro ilmiy-nazariy jurnal. – 2022. – №4. – B. 83–88.
6. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. Applied Sciences, 11(19), 9117.
7. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).

O'ZBEK BOLALAR NUTQINING KORPUSGA ASOSLANGAN PSIXOLINGVISTIK TADQIQI: CHILDES TIZIMI ASOSIDA YONDASHUV

Abduraxmonova Sohiba Aktam qizi

Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti,
Jurnalistika va o'zbek filologiyasi fakulteti, tayanch doktoranti

E-mail: abduraxmonovasohiba@gmail.com

Annotatsiya: Mazkur maqolada o'zbek bolalar nutqining psixolingvistik jihatdan korpusga asoslangan tahlili CHILDES tizimi asosida ko'rib chiqiladi. Maqolada bolaning tilni o'zlashtirish bosqichlari, grammatik birliklarning shakllanishi, kommunikativ kompetensiyaning rivoji korpus lingvistikasi va psixolingvistika yondashuvi asosida yoritiladi. CHILDES tizimidan foydalanish orqali o'zbek tilidagi bola nutqini standartlashtirilgan formatda kodlash imkoniyati tahlil qilinadi. Shuningdek, maqolada O'zbek bolalar nutqi korpusini yaratishda yuzaga keladigan yutuqlar muhokama qilinadi.

Kalit so'zlar: bola nutqi, psixolingvistika, til o'zlashtirish, til korpusi, CHILDES

Abstract:In this article, a corpus-based psycholinguistic analysis of uzbek children's speech is considered based on the CHILDES system. The article covers the stages of language acquisition of a child, the formation of grammatical units, the development of communicative competence based on the approach of corpus linguistics and psycholinguistics. Using the CHILDES system, the possibility of coding children's speech in the Uzbek language in a standardized format is analyzed. Also, the article discusses the achievements in the creation of the corpus of Uzbek children's speech.

Keywords:speech child, psycholinguistics, language learning, language corpus, CHILDES

Аннотация:В данной статье рассматривается корпусный психолингвистический анализ речи узбекских детей на основе системы CHILDES. В статье на основе подходов корпусной лингвистики и психолингвистики рассматриваются этапы овладения языком ребенка, формирования грамматических единиц, развития коммуникативной компетентности. С помощью системы CHILDES анализируется возможность кодирования детской речи на узбекском языке в стандартизированный формат. Также в статье рассматриваются достижения в создании корпуса речи узбекских детей.

Ключевые слова: речь ребенка, психолингвистика, овладение языком, языковой корпус, чилдес

Bolalar nutqining shakllanishi va rivojlanishini o'rganish psixolingvistikaning asosiy yo'nalishlaridan biri hisoblanadi. Til nafaqat ijtimoiy aloqa vositasi, balki tafakkur, idrok va kognitiv jarayonlar bilan chambarchas bog'liq psixik tizimdir. Bola til o'zlashtirar ekan, u tilni grammatik struktura sifatida emas, balki voqelikni anglash va ifodalash vositasi sifatida idrok qiladi. Shu sababli, bolalar nutqini tahlil qilish orqali nafaqat lingvistik, balki psixologik va kognitiv jarayonlarni ham chuqurroq tushunish mumkin. So'nggi yillarda tilni empirik asosda o'rganish yo'nalishida korpus lingvistikasi keng qo'llanila boshladi. Korpus lingvistikasi – tabiiy til namunalarini raqamli shaklda to'plab, ular ustida statistik va strukturaviy tahlil olib borish imkonini beruvchi zamonaviy yo'nalishdir. Ushbu maqola O'zbekiston sharoitida ilk bor o'zbek bolalarning tabiiy og'zaki nutq namunalaridan tuzilgan korpus asosida ularning til o'zlashtirishdagi psixolingvistik xususiyatlarini tahlil qilishga qaratiladi.

Bolalar nutqi va uni o'rganish muammosi xalqaro psixolingvistika sohasida uzoq yillardan buyon keng muhokama qilinmoqda. L.S. Vygotskyning "Til va tafakkur" konsepsiyasi[1], J. Piagetning bola tafakkurining bosqichma bosqich rivojlanishiga oid nazariyasi[2], D. Slobin[3] va M. Tomasellolarning[4] til o'zlashtirish jarayoniga oid yondashuvlari bu sohada asosiy nazariy

tayanchlar hisoblanadi. Xususan, D.Slobin tomonidan ilgari surilgan “o‘zlashtirishga tayyor til” (language-ready brain) g‘oyasi bolalar tilini tahlil qilishda kognitiv asoslarni yoritadi.

O‘zbek tilshunosligida bolalar nutqi bo‘yicha ham qator ilmiy ishlar yaratilgan. Jumladan, O‘zbek tilshunosligida O.Umarxo‘jayeva[5] N.Sayidiraximova[6], G.Iskandarova[7], M.Qurbonova[8], va O.Saidaxmedovalar[9] tomonidan bolalar nutqini qator tadqiqotlarda kuzatilgan va o‘zbek tilining birinchi til sifatida fonetik, grammatik xususiyatlari lingvistik, sotsiolingvistik, pragmalingvistik va psixolingvistik jihatdan tadqiq etilgan.

Til o‘zlashtirish inson psixikasining murakkab, ko‘p bosqichli va ko‘p omilli jarayonidir. Bola nutqini o‘rganish psixolingvistika doirasida til va tafakkur, til va kognitiv rivojlanish, til va kommunikatsiya munosabatlarini o‘z ichiga oladi. Mazkur tadqiqot o‘zbek bolalarning tabiiy nutq namunalari asosida empirik tahlil qilishni maqsad qilib qo‘yadi. Bu jarayonda xalqaro CHILDES[10] tizimidan namuna sifatida foydalanish o‘zbek tilshunosligida yangi yo‘nalish – korpusga asoslangan psixolingvistik tadqiqotlar uchun poydevor bo‘la oladi. CHILDES (Child Language Data Exchange System) – bolalar nutqini kodlash, transkripsiya qilish va tahlil etish uchun ishlab chiqilgan xalqaro tizimdir. Ushbu tizimning asosiy komponentlari CHAT (transkripsiya formati) va CLAN (tahlil vositalari) bo‘lib, ular til o‘zlashtirish jarayonini tizimli o‘rganish imkonini beradi. Ma‘lumotlar o‘zaro taqqoslanadigan, strukturalangan va statistik jihatdan tahlil etilishi mumkin bo‘lgan shaklda saqlanadi. CHILDES loyihasi 1984-yilda Brian MacWhinney tomonidan yaratilgan bo‘lib, u bolalar nutqining audio va video yozuvlarini transkripsiya qilish, kodlash va tahlil qilishga mo‘ljallangan. CHILDES platformasi CHAT formati, CLAN dasturi va TalkBank bazasi orqali tilshunoslar uchun boy resurslar yaratadi. Ushbu model hozirda 30 dan ortiq tillarda bolalar nutqiga oid korpuslarni o‘z ichiga oladi.

O‘zbek bolalar nutqining CHILDES modeliga mos korpusini yaratish quyidagi bosqichlardan iborat bo‘ladi:

- Nutq namunalarini yig‘ish (intervyu, uy sharoitida yozuv, bolalar bog‘chalari);
- Transkripsiya qilish (CHAT formati asosida);
- Morfosintaktik belgilash (fe‘llar, otlar, gap tuzilmalari, qo‘shimchalar);
- Psixolingvistik holatlarni belgilash (emotsional reaksiyalar, pauzalar, takrorlar, diskursiv markerlar);

–Tahlil qilish (CLAN dasturi orqali statistik va funksional tahlillar).

Namuna: O‘zbekcha transkripsiya modeli

*CHI: bu meniki.

*MOT: yo‘q, bu akangniki.

%mor: bu/dem.pro meniki/adj.poss

@End

Amaliy ahamiyati

O‘zbek tilida korpusga asoslangan psixolingvistik tadqiqotlar natijasida:

- til o‘zlashtirish mexanizmlari empirik asosda o‘rganiladi;
- nutq nuqsonlari va ularni tuzatish metodlari ishlab chiqiladi;
- sun‘iy intellekt va avtomatik tahlil vositalarida o‘zbek tilining imkoniyatlari kengayadi;
- xalqaro lingvistik tadqiqotlarda o‘zbek tili uchun ma‘lumotlar bazasi shakllanadi.

Mazkur maqolada o‘zbek bolalar nutqini psixolingvistik jihatdan tahlil qilishda korpusga asoslangan yondashuv, xususan CHILDES tizimining imkoniyatlari tahlil qilindi. O‘zbek tilida bunday korpusni yaratish tilshunoslik va ta‘lim sohalarida muhim manba bo‘lishi mumkin. Tadqiqot natijalari bola nutqining rivojlanishini aniq kuzatish, diagnostika va tuzatish tizimlarini ishlab chiqishda qo‘llanilishi mumkin.

Foydalanilgan adabiyotlar ro‘yxati:

1. Vygotsky, L. S. (1986). Thought and Language. Cambridge: MIT Press.
2. Piaget, J. (1959). The Language and Thought of the Child. London: Routledge.

3. Slobin, D. I. (1985). The Crosslinguistic Study of Language Acquisition. Hillsdale, NJ: Erlbaum.
4. Tomasello, M. (2003). Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge: Harvard University Press.
5. Умархўжаева О. Болалар нутқининг баъзи фонетик хусусиятлари Ўзбек тили ва адабиёти. – Тошкент, 1997. – № 5. – Б. 52-53
6. Сайидираҳимова Н. Мактабгача ёшдаги ўзбек болалар нутқининг лингвистик хусусиятлари: Филол. фан. номз. ...дисс. – Тошкент, 2004. – 126 б
7. Искандарова Г. Болалар нутқига оид баъзи мулоҳазалар//Ўзбек филологиясига оид тадқиқотлар: Илмий-амалий анжуманлар материаллари . . – Тошкент: ЎЗМУ, 2009 – Б.200-203.
8. Курбонова М. Ўзбек болалар нутқи лексикасининг социопсихолингвистик тадқиқи: Филол. фан. номз. ... дисс. – Тошкент, 2009. – 154 б.
9. Saidaxmedova O. Tabiiy tartib gepotezasining o‘zbek tili uchun talqini. Toshkent – 2021
10. MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. Mahwah, NJ: Erlbaum.

O'ZBEK TILIDAGI SO'Z BIRIKMALARINI FORMALLASHTIRISH BOSQICHLARI

Jo'raqulova Madina Subhonovna

O'zbekiston Milliy universiteti
Kompyuter lingvistikasi
va amaliy tilshunoslik kafedrası o'qituvchisi
E-mail: joraqulova_m@nuu.uz

Qo'ziyeva Sitora Ibodulla qizi

O'zbekiston Milliy universiteti 4-bosqich talabasi
E-mail: nkosmos1199@gmail.com

Annotatsiya: Bu maqolada so'z birikmalarining formal modellari ifodalangan bo'lib, kompyuter lingvistikasi fanining avtomatik tarjima, avtomatik sintaktik tahlil va NLP – tabiiy tilning qayta ishlanishi yo'nalishlarida lingvistik baza sifatida foydalanish mumkin. O'zbek tilshunosligida amalga oshirilgan so'nggi tadqiqotlar natijasida yangi yo'nalishlarning paydo bo'lishi, til birliklarining modellari aniqlanishi, lingvistik amallarni, tahlillarni avtomatik bajarish ishlari jadal rivojlandi. Til korpuslarining qurilishi, morfonalizatorning ishlab chiqilishi, keyingi qadam sifatida esa sintaktik analizator dasturining lingvistik va dasturiy ta'minoti ishlab chiqilishi muhim vazifa hisoblanadi. Aynan mana shunday avtomatik tahlil dasturlari uchun til birliklarining modellari ishlab chiqilishi zarur. Mazkur maqolamizda o'zbek tilidagi so'z birikmalarining so'zlar o'rtasidagi grammatik aloqani ta'minlovchi vositalarning ishtirokiga ko'ra turlari aniqlandi va lingvistik modellar taklif qilindi.

Kalit so'zlar: sintaktik analizator, korpus, so'zlarning birikuv usullari, analitik va sintetik so'z birikmalari strukturasi, modellashtirish, avtomatik tahlil, formal model

Annotation. This article proposes formal models of Uzbek word combinations, aimed at supporting key areas of computational linguistics such as machine translation, syntactic parsing, and natural language processing (NLP). The study classifies phrase types based on grammatical connectors between words and offers corresponding linguistic models. These models are intended to contribute to the development of tools such as morphoanalyzers and syntactic analyzers, and to enhance automated linguistic analysis.

Keywords. syntactic analyzer, corpus, word combinations, analytic and synthetic constructions, modeling, automatic analysis, formal model

Аннотация. В данной статье предлагаются формальные модели узбекских словосочетаний, направленные на поддержку ключевых областей компьютерной лингвистики, таких как машинный перевод, синтаксический анализ и обработка естественного языка (NLP). В исследовании классифицируются типы словосочетаний на основе грамматических связей между словами и предлагаются соответствующие лингвистические модели. Эти модели предназначены для развития таких инструментов, как морфоанализаторы и синтаксические анализаторы, а также для совершенствования автоматизированного лингвистического анализа.

Ключевые слова: синтаксический анализатор, корпус, словосочетания, аналитические и синтетические конструкции, моделирование, автоматический анализ, формальная модель

Kirish. O'zbek tilini kompyuterlashtirish sohasidagi fundamental va amaliy tadqiqotlar natijalari tabiiy til xususiyatlari bilan bog'liq bo'lib, bugungi kunda o'zbek tilining grammatik (morfoloqik va sintaktik) qoidalari asosida matematik modellari va ularning aksiomatik nazariyasini yaratish kompyuter lingvistikasi sohasidagi dolzarb vazifalardan biridir. Grammatik qoidalarning

ontologik modellarini aniqlash orqali o'zbek tilining kompyuter tahlil qila oladigan dasturiy ta'minot uchun matematik mantiqqa asoslangan formal grammatikasi yaratilishi zarur. Zotan sun'iy intellekt texnologiyalari yordamida amalga oshiriladigan har qanday til texnologiyasiga tegishli yangiliklar muayyan qoidalarga bo'ysungan holda amalga oshirilishi lozim.

Keyingi yillarda ko'plab tillar uchun sintaktik analizatorlar yaratish talabi sun'iy intellekt va tabiiy tilni qayta ishlash (NLP – Natural Language Processing) texnologiyalarining jadal rivojlanishi bilan bog'liq. O'zbek tilida ham bunday tizimlarga ehtiyoj tobora ortib bormoqda, avtomatik tarjima, matn tahlili va ovozli interfeyslar kabi yo'nalishlarda bu, ayniqsa, seziladi. Matnning grammatik tuzilishini aniqlovchi sintaktik analizatorlar samarali ishlashi uchun so'zlar o'rtasidagi sintaktik bog'liqliklarni to'g'ri modellashtirish talab qilinadi. Sintaktik analizatorlar tabiiy tilda yaratilgan matnlarni struktura jihatdan tekshiradi, ularning grammatik qoidalar asosida shakllanganligini aniqlaydi va ma'lumotlarni formal modellashtirish imkoniyatini yaratadi. Sintaktik analizatorlar avtomatik tarjima, sun'iy intellekt, chatbotlar uchun matnni qayta ishlash, programmashtirish tillari uchun kod sintaksisini tekshirish jarayonlarida qo'llanilishi bilan ahamiyatlidir.

Materiallar va metodlar. O'zbek tili sistem-struktural tilshunosligida sintaksis masalalari uch yo'nalishda o'rganildi:

1. Formal yo'nalish;
2. Substansial (formal-funksional) yo'nalish;
3. Mazmuniy sintaksis yo'nalishi

Formal sintaksis bevosita kuzatishda berilgan, nutqiy sintaktik hodisalarni o'rgansa, substansial (formal-funksional) sintaksis esa asosiy diqqat-e'tiborini til hodisalarining ichki (bevosita kuzatishda berilmagan tomoniga qaratadi. Ular bir-birini ham taqozo, ham inkor etadi. Formal sintaksis Mutlaq hodisalar bilan, substansial sintaksis esa Mutlaq mohiyat bilan shug'ullanadi. Shu sababli, ularning vazifalarini mutlaq qarama-qarshi qo'yish mumkin emas. Substansial sintaksis tadqiqotlari uchun formal sintaksis tomonidan kiritilgan yutuqlar zamin, poydevor vazifasini o'taydi. Morfologik ko'rsatkichsiz va morfologik ko'rsatkichli so'z birikmasining turini aniqlashda formal-funksional tilshunoslik vakillari tomonidan yaratilgan so'z birikmalarining lisoniy-sintaktik qoliplari asosida tuzilgan modellardan foydalanish sintaktik teglar tizimining mukammal bo'lishiga olib keladi.

Modellashtirish jarayonida til sathlarining bir-biriga bog'liq ekanligini e'tiborga olgan holda, ularning o'zaro munosabati hamda har bir sathni o'ziga xos xususiyatlarini o'rganish ahamiyatlidir. R.Tillayeva ilmiy tadqiqot ishida ham boshqa til birliklari bilan bir qatorda, sintaktik birliklarni modellashtirish masalasini ham atroflicha o'rganib, o'z fikr-mulohazalarini bildiradi: Til sistemasining elementlari o'ziga yaqin turgan yuqori sath birliklari tarkibida funksiyalashadi. Shuning uchun ma'lum sath birligining mohiyati haqida ma'lumot olish uchun bu birlikning yuqori sathdagi funksiyasini o'rganishga ham e'tibor berish kerak bo'ladi. Mana shu nuqtayi nazardan yondashganda, sintaktik sath til sistemasining yuqori sathi sifatida o'z ichida barcha birliklarning funksiyasini jamlaydi. Gap va uning qismlarida til birliklarining barchasi jamlangan bo'ladi [Tillayeva, 2016]. Sintaktik birliklarni modellashtirish deganda, asosan, so'z birikmasi va gapni modellashtirish nazarda tutiladi. Bir necha maqolalarda, tadqiqot ishlarida so'z birikmasining o'ziga xos jihatlarini e'tiborga olgan holda, turlicha qarashlar mavjudligini kuzatamiz. S.Boysinov, O.Mamirovlar "Kompyuter lingvistikasida modellashtirish metodidan foydalanish" maqolasida til birliklarini modellashtirishda ingliz tili materiallari namuna sifatida keltiriladi. So'z birikmasi modeli esa uning umumiy xususiyatidan kelib chiqib, quyidagicha izohlanadi: T- H [tobe-hokim] ushbu modelni keltirish bilan cheklangan [Boysinov, Mamirov, www.jdpu.uz]. So'z birikmasini modellashtirishda o'zbek tilining xususiyatlarini o'zida aks ettiruvchi belgilar orqali modellashtirish yuqori natijalarni beradi. Sintaktik analizda faol ishtirok etuvchi birikuv usullari tilshunoslik va kompyuter lingvistikasi nuqtayi nazaridan bir-biridan biroz farq qilishi mumkin. Shunga ko'ra, so'z birikmalari bir necha xususiyatlariga ko'ra tasnif qilinadi. Grammatik aspektda so'z birikmalarining shakli degan tushuncha mustaqil so'zlar o'rtasidagi

o‘zaro sintaktik munosabatlarni ifodalash usullari - bitishuv, boshqaruv, moslashuv usullarini o‘z ichiga oladi. So‘zlar o‘rtasidagi grammatik aloqani ta‘minlovchi vositalarga ko‘ra birikuv usullarini quyidagilarga ajratish mumkin: sintetik usul, analitik usul va so‘z tartibi.

Natija. So‘z birikmalari modellaridan kelib chiqib aytishimiz mumkinki, o‘zbek tilida so‘zlar, asosan, sintetik usul yordamida, – ya‘ni grammatik qo‘shimchalar orqali bog‘lanadi. Bu usulning tub mohiyati shundan iboratki, so‘z birikmasini hosil qilgan komponentlardan biri biror qo‘shimchani oladi, shu affiks yordamida bir so‘z ikkinchi so‘zga grammatik jihatdan bog‘lanadi va fikr almashish uchun tayyor material vazifasini o‘taydi. Bunday bog‘lanish so‘z o‘zgartiruvchi va shakl yasovchi deb atalgan qo‘shimchalar yordamida yuzaga keladi. Masalan, “*maktabning o‘quvchilari*” so‘z birikmasida *qaratqich aniqlovchi va egalik qo‘shimchalari* so‘zlar o‘rtasidagi bog‘liqlikni vujudga keltirgan. Sintetik bog‘lanish lug‘aviy va sintaktik shakl yasovchi deb atalgan qo‘shimchalar yordamida yuzaga keladi. Bunda mustaqil so‘zlar *egalik affikslari, kelishik qo‘shimchalari, sifatdosh, ravishdosh qo‘shimchalarini* qabul qiladi. Undan tashqari, lug‘aviy-sintaktik qo‘shimchalar: **-dek/-day, -dagi, -niki** kabilarni ham qabul qilgan so‘zlarni o‘z ichiga oladi. Ba‘zi so‘z yasovchi affikslar ham, masalan: -li, -lik, -dagi, -day kabi yasovchilar ham so‘zlarning o‘zaro aloqaga kirishuvida ishtirok etadi. Bu va boshqa juda ko‘p formal elementlar tilning butun morfologik strukturasini qamrab olgan. Ayrim qo‘shimchalarning ish ko‘rish doirasi bir qadar cheklangan, masalan, egalik va kelishik qo‘shimchalari faqat ot kategoriyasidagi so‘zlarga xos.

Quyida sintetik usulda hosil bo‘ladigan so‘z birikmalari qoliplari va shu qoliplar asosida shakllantirilishi mumkin bo‘lgan modellar keltiriladi;

[ot^{tushum kelishigi} → fe‘l] => matnlarni tahlil qilish; matnlarni tahrirlash;

[ot^{o‘rin-payt kelishigi} → fe‘l] => bunda o‘rganmoq; bartaraf etilgandagina yondashilgan bo‘lmoq; o‘rinda hosil qilish;

[ot^{jo‘nalish kelishigi} → fe‘l] => bir-biriga mos kelmoq; masalaga yondashilmoq; o‘zgarishiga sabab bo‘lmoq; fikrlarga e‘tibor qaratmoq;

[ot^{chiqish kelishigi} → fe‘l] => umringdan topgan kabi qoliplar asosida ism+fe‘l ko‘rinishidagi so‘z birikmalari uchun [Sub^{Cs3} → V]; [Sub^{Cs4} → V]; [Sub^{Cs5} → V]; [Sub^{Cs6} → V] kabi modellar mos keladi. Ismning turli ko‘rinishlariga ko‘ra turlari (ot, olmosh, sifat, son, harakat nomi, sifatdosh) bo‘yicha [N^{Cs3} → V]; [Adj^{Cs3} → V]; [Num^{Cs3} → V]; [Pr^{Cs3} → V]; [Inf^{Cs3} → V]; [Gerund|Past participle → V] kabi modellar so‘z birikmalarini farqlashga yordam beradi. Bunda Cs³ tushum, Cs⁴ jo‘nalish, Cs⁵ o‘rin-payt, Cs⁶ chiqish kelishigini; **Sub** ismni, **V** fe‘lni ifodalovchi teg hisoblanadi.

Shuningdek, ism+ism ko‘rinishida ham so‘zlar bir-biri bilan sintetik usulda birikma hosil qilishi bo‘lishi mumkin. Masalan:

[Ism^{qaratqich kelishigi} → ism^{egalik qo‘shimchasi}] => daraxtning mo‘rti. Mazkur qolipga quyidagi so‘z birikmalari modellari mos keladi:

- 1) [N^{Cs} → N^{Pos}]: tovarning belgisi;
- 2) [N^{Cs} → Adj^{Pos}]: qog‘ozning qattig‘i;
- 3) [Adj^{Cs} → N^{Pos}]: kattasining imkoniyati;
- 4) [Adj^{Cs} → Adj^{Pos}]: olmaning/kattasining chuchugi;
- 5) [N^{turdosh Cs} → Adj^{Pos}]: qizlarning biri;
- 6) [Num^{Cs} → Num^{Pos}]: oltmishning choragi;
- 7) [N^{Cs} → Ger^{Pos}]: elementlarning ifodalanishi;
- 8) [Inf^{Cs} → N^{Pos}]: uyalishning o‘rni;
- 9) [Inf^{Cs} → Ger^{Pos}]: yemoqning qusmog‘i;
- 10) [N^{Cs} → GerPP^{Pos}]: yurakning to‘xtagani;
- 11) [Ger^{Cs} → N^{Pos}]: buyurganning mahsuloti;
- 12) [N^{Cs} → Prv^{Pos}]: talabalikning buguni;
- 13) [Pr^{Cs} → N^{Pos}]: bizning nutqimiz;
- 14) [Prv^{Cs} → N^{Pos}]: hozirning huzuri.

Bunda ot = N, sifat = Adj, son = Num, turdosh ot = N^{sub}, olmosh = Pr, ravish = Prv, harakat nomi = Infinitive, sifatdosh = Gerund teglari bilan; qaratqich kelishigi = Case (yoki Cs), egalik qo'shimchasi = Possessive (yoki Pos) teglari bilan belgilandi.

Ism+ism qolipli so'z birikmalarida qaratqich+qaralmish munosabati bo'lsa, a'zolaridan biri belgisiz qo'llangan bo'lsa ham bu so'z birikmasini moslashuv munosabatli so'z birikmasi sifatida modellashtirishimiz mumkin.

$[N \rightarrow N^{Pos}] \Rightarrow$ baxt manzili

Biroq xuddi shu qolipga mos keluvchi izohlovchi+izohlanmish munosabatli **Davlat qo'mitasi, Navoiy bog'i** kabi so'z birikmalarining modellarini yuqoridagi modellardan farqlash zarur. Chunki bu yerda moslashuv munosabatli so'z birikmasi emas, bitishuv munosabatli so'z birikmasi turi vujudga kelgan.

Shuningdek, sintetik usulda yasalgan so'z birikmalari quyidagi usullarda ham ham voqelanishi mumkin:

$[fe' I^{sifatdosh} \rightarrow ot] \Rightarrow [V^{ger} \rightarrow N]$: topgan quvonch; topgan mamnuniyat

$[fe' I^{ravishdosh} \rightarrow fe' I] \Rightarrow [V^{ger} \rightarrow V]$: kulib gapirmoq;

$[ot^{dagi} \rightarrow ot] \Rightarrow [(N+dagi) \rightarrow N]$: asarlaridagi go'zallik, asarlaridagi nafosat; quyidagi jihatlar; ko'rinishidagi birikmalar.

Analitik usul esa yordamchi so'zlar (bilan, uchun, kabi), bog'lovchilar, so'z tartibi va har qanday bog'lama vazifasida keluvchi so'zlar (masalan: bor, yo'q so'zlari *ilmi yo'q inson, o'qituvchi bo'lib faoliyat ko'rsatmoq* birikmalaridagi *yo'q, bo'lib* so'zlarining vazifasi ahamiyatli) orqali yuzaga chiqadi. Modellashtirish nuqtayi nazaridan sintetik usuldagi bog'lanishlar dasturiy jihatdan osonroq aniqlanadi, chunki ular aniq morfologik belgilarga ega. Analitik usullar esa kontekstual aniqlikni talab qiladi, bu esa statistik yondashuvlarga ehtiyoj tug'diradi. Analitik usulda birikmani tashkil qilgan komponentlar, asosan, hech qanday qo'shimcha olmaydi, agar qo'shimcha olsa ham, bu morfologik vosita ikkilamchi pozitsiyada, ya'ni lug'aviy ma'noga ta'sir etmagan holda o'ziga xos grammatik ma'noni ta'minlaydi (*kitoblar olib kelindi*).

So'zlarning ushbu o'zaro birikish usullari boshqa tillarda ham mavjud bo'lib, kompyuter lingvistikasida matnlarni avtomatik tahlil qilishda har ikkala usuldan unumli foydalaniladi. Masalan, lotin tili sintetik tillar sirasiga kirib, u morfemalarni o'ziga to'g'ridan to'g'ri "yopishtiradi". Analitik usul esa o'z xususiyatini predlogli tillarda ko'rsatadi (*masalan, the leaves "of" a tree*) o'zbek tili esa har ikkala usulning kombinatsiyasiga ega. Qaraymiz:

1. $[(Ot+uchun) \rightarrow fe' I] \Rightarrow [Noun+A_1 \rightarrow Verb]$: ta'minot uchun shakllantirilmoq;

2. $[(Ot+kabi) \rightarrow fe' I] \Rightarrow [Noun+A_1 \rightarrow Verb]$: umumiylik kabi jihatlar

Muhokama. O'zbek tili korpusi asosida yaratilayotgan sintaktik analizatorlar uchun so'z birikuv usullarining to'g'ri modellashtirilishi nihoyatda muhim. Sintetik usulni modellashtirishda mavjud morfologik tahlil vositalaridan samarali foydalanish mumkin, chunki bu usul grammatik qo'shimchalarga tayanadi. Biroq, analitik usullarni to'g'ri aniqlash uchun semantik kontekst va nutq birligi chegaralarini aniqlash lozim, bu esa chuqur o'rganishni talab qiladi.

Dependency Parsing jarayonida "bosh so'z" va boshqa barcha so'zlar o'rtasidagi munosabatlarni aniq belgilaydigan tobelik daraxti (Dependency Tree) deb nomlanuvchi grammatik struktura shakllantiriladi. Tobe munosabatlar tahlili (Dependency Parsing) tobelik grammatikasiga asoslangan bo'lsa, tarkibiy qismlar tahlili (Constituency Parsing) erkin kontekstli grammatikaga (context-free grammar) asoslanadi. Sintaktik tahlilning bu turi matn ma'lumotlaridagi so'z birikmalari ustida amallar bajaradi. Tarkibiy qismlar tahlili jarayonida matn grammatik toifaga asoslanib, quyi so'z birikmalari, tarkibiy qismlarga ajratiladi. Grammatik birliklariga otli birikma (NP), fe'lli birikma (VP) yoki ba'zan old so'z birikmasi (PP) misol bo'ladi. Sintaktik daraxtlar tabiiy tilda gap yoki so'z birikmasining sintaktik tuzilishi grafik tasvirini ifodalaydi.

Hozirda dunyo tillarida keng qo'llanilayotgan dependency-parsing yondashuvi O'zbek tilida ham qo'llanilishi mumkin. Bunda so'zlar o'rtasidagi grammatik bog'liqliklar tarmoqli

ko‘rinishida tasvirlanadi. Korpusdagi real matnlar asosida tahlil olib borish orqali ushbu yondashuvni moslashtirish mumkin.

Xulosa. O‘zbek tilidagi so‘z birikmalarini to‘g‘ri modellashtirish – sintaktik analizatorlar sifatini oshirishning asosi hisoblanadi. Nafaqat so‘z birikmasi hosil bo‘lishining sintetik va analitik usullarini chuqur tahlil qilish, balki sintetik-analitik hamda so‘z tartibi usullarini ham dasturiy jihatdan modellashtirish uchun maxsus korpuslar yaratish va zamonaviy NLP yondashuvlaridan foydalanish – istiqbolli yo‘nalishdir. Bu boradagi izlanishlar nafaqat tilshunoslik, balki sun‘iy intellekt sohasida ham katta ahamiyat kasb etadi.

Foydalanilgan adabiyotlar:

1. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). UZBEK ELECTRONIC CORPUS AS A TOOL FOR LINGUISTIC ANALYSIS. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
2. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
3. Rahmatullaev Sh. O‘zbek frazeologiyasining ba’zi masalalari. (Frazeologik polisemiya, sinonimiya, antonimiya va omonimlik): Filol.f.doktr...diss. – Toshkent: O‘qituvchi, 1966.- 264 b
4. Bushuy T., Safarov Sh. Til qurilishi tahlil metodlari va metodologiyasi. – Toshkent, 2007-yil.
5. Sayfullayeva R., Mengliyev B. va b. Hozirgi o‘zbek adabiy tili. – Toshkent, 2009.
6. Abdullayev F., Ibrohimova F. O‘zbek tilida boshqaruv. Monografiya. – T., 1962-yil. O‘zbek tili grammatikasi II tom. -T., 1978.
7. Rahimov A. Kompyuter lingvistikasi asoslari. – T., 2011.
8. Tillayeva R. Lisoniy birliklarni modellashtirish. Magistrlik dissertatsiyasi ishi. – Namangan, 2016.
9. Boysinov S., Mamirov O. Kompyuter lingvistikasida modellashtirish metodidan foydalanish. Zamonaviy dunyoda innovatsion tadqiqotlar: Nazariya va amaliyot nomli ilmiy, masofaviy, onlayn konferensiya materiallari.
<https://doi.org/10.5281/zenodo.7515431>

O'ZBEKISTONDA KOMPYUTER LEKSIKOGRAFIYASI YUZASIDAN OLIB BORILGAN TADQIQOTLAR

Mengziyotova Faridabonu

O'zbekiston Milliy universiteti
Jurnalistika va o'zbek filologiyasi fakulteti
2-kurs 201-guruh talabasi
E-mail: mengziyotovafarida@gmail.com

Annotatsiya: Mazkur maqolada kompyuter leksikografiyasi yuzasidan amalga oshirilgan tadqiqotlar tahlil qilingan. Hozirgi kunda zamonaviy texnologiyalar taraqqiyoti va tilshunoslik bilan aloqasi natijasida lug'at tuzish jarayonlari yangi bosqichga olib chiqildi va inson mehnatini yengillashtirish, ma'lumotlarni tartibga solish imkoniyatlari yanada kengaytiruvchi lug'at bazalari tahlil qilingan. Xususan, elektron lug'atlar, WordNet, tezaurus, ontologiyalar yaratish bo'yicha olib borilgan ishlar ma'lum ma'noda o'z ifodasini topgan.

Kalit so'zlar: amaliy leksikografiya, axborot qidirish tezauruslari, ontologik lug'atlar, lingvistik ontologiyalar, WordNet tipidagi lug'atlar

Annotation: This article analyzes research conducted in the field of computer lexicography. Nowadays, due to the development of modern technologies and their relationship with linguistics, the process of dictionary compilation has entered a new stage. Lexical databases that facilitate human labor and organize information more effectively have been examined. In particular, works related to the creation of electronic dictionaries, WordNet, thesauri, and ontologies are partially reflected.

Keywords: applied lexicography, information retrieval thesauri, ontological dictionaries, linguistic ontologies, WordNet-type dictionaries.

Аннотация: В данной статье проанализированы исследования, проведённые в области компьютерной лексикографии. В настоящее время благодаря развитию современных технологий и их связи с лингвистикой процессы составления словарей вышли на новый этап. Рассмотрены лексикографические базы данных, облегчающие труд человека и упорядочивающие информацию. В частности, частично отражены работы, связанные с созданием электронных словарей, WordNet, тезаурусов и онтологий.

Ключевые слова: прикладная лексикография, тезаурусы для информационного поиска, онтологические словари, лингвистические онтологии, словари типа WordNet.

Bugungi kunda kompyuter lingvistikasi va uning asosiy yo'nalishlarida arzigulik tadqiqotlar amalga oshirilgan, xususan, kompyuter leksikografiyasi an'anaviy leksikografiyadan farqli ravishda turli lug'at va ma'lumot bazalarini avtomatlashtirish, tahlil qilish va ulardan samarali foydalanish imkoniyatlarini taqdim etadi. Lug'atshunoslikka kompyuter texnologiyalarining tadbiq etilishi natijasida qator ko'p funksiyali zamonaviy lug'atlar yaratildi. Jumladan Wordnet, tezaurus, ontologiya kabi elektron lug'atlarni yaratish yuzasidan amalga oshirilgan tadqiqotlar bu yo'nalishning taraqqiy etishida asosiy omil hisoblanadi.

O'zbekistonda kompyuter leksikografiyasi sohasida bir qator olimlar tomonidan samarali tadqiqotlar olib borilgan. Jumladan, A.Rahimov o'zbek tilining kompyuter lingvistikasi sohasidagi asosiy tushunchalari va metodlarini yoritib bergan. S.Muhamedova esa til korpuslari va ularning yaratilishi, elektron lug'atlar va ularning yaratilishi bo'yicha ilmiy izlanishlar olib borgan. Y. Baxriddinova ingliz va o'zbek tillarida kompyuter lingvistikasi leksikografik muammolari yuzasidan tadqiqotlar olib borgan. U ingliz tilidagi tilshunoslikka oid zamonaviy kompyuter atamalarining o'zbek tiliga moslashuvi, ularni tizimli ravishda lug'at shaklida taqdim etish

masalasini tadqiq qiladi. Uning fikricha, elektron lugʻatlarni foydalanuvchiga qulay interfeysda ishlab chiqish til madaniyatiga xizmat qiladi.

Lugʻatshunoslikning asosiy oʻrganish obyekti – bu soʻzning barcha xususiyatlari va munosabatlaridir. Hozirgi paytda leksikografiya nafaqat amaliy, balki axborot – kognitiv va metodologik muammolarini hal qilish bilan oʻz rolini oshirmoqda. Leksikografiya sohasiga kompyuter texnologiyalarining tatbiq etilishi lugʻat jarayonlarini ancha yengillashtiradi. Lugʻat tuzish jarayoninig naqadar murakkab ekanligini bir tilshunos olim hazilomuz tarzda shunday deydi: „Agar biror kishi ogʻir jinoyat sodir qilsa, uni surgun qilish shart emas. U lugʻat tuzsin, shunda u katoraning barcha azoblarini, qiyinchiliklarini tortgan boʻladi“. Ijtimoiy tafakkurni yuksaltirishda lugʻatlarning roli beqiyosligi jahon tajribasida oʻzini toʻla oqlagan haqiqatdir. Insonning mustaqil fikrlashi–yu, soʻzlarni aniq va ravon tushunishi yoshligidan boshlanadi.

Kompyuterda yaratilgan birinchi matnlar korpusi Braun korpusi hisoblanadi, u 1961–yilda Braun universitetida yaratilgan, har biri 2000 soʻzli 500 ta matn fragmentini oʻz ichiga oladi. 1970–yilda 1mln soʻzni oʻz ichiga olgan matnlar korpusi asosida rus tilining chastotali lugʻati yaratildi. 1980–yilda Shvetsiyaning Upsala universitetida ham rus tili matnlar korpusi yaratildi. Keyinchalik kompyuter leksikografiyasining rivojlanishi natijasida katta hajmli matnlar korpusiga ehtiyoj tugʻildi. Yaʼni 1mln ta soʻz elektron lugʻatlar bazasi uchun yetarli emas. Shu sababli yirik hajmli matnlar korpusi yaratila boshlandi.

Masalan, rus tilining milliy korpusi hajmi hozirgi kunda 149 mln soʻzdan iborat. Leksikografik maʼlumotlar bazasining sifati faqat manbalarning hajmi va xilma-xilligi bilan emas, balki leksikografik xizmatning rivojlanishi bilan ham belgilanadi. 1988–yilda Buyuk Oksford lugʻatining yangi 20 jildlik nashri yakunlandi. Lugʻatning butun korpusi maʼlumotlar bazasi shaklida kompyuterda saqlanadi. Shu bilan birga mashinaning ishlash tezligi odam tezligidan bir necha baravar katta. Bu ham, albatta, qoʻl mehnatini yengillashtirishda samaralidir. Lugʻatlarning turlari haqida gapirganda birinchi boʻlib elektron lugʻatlarni yodga olish kifoya. Elektron lugʻat (kompyuter lugʻati) – maxsus kodlangan lugʻat yozuvlarini oʻz ichiga olgan, koʻpincha morfologik shakllarni hisobga olgan holda va soʻz birikmalarini qidirish, shuningdek, tarjima yoʻnalishini oʻzgartirish qobiliyatiga ega boʻlgan kerakli soʻzlarni tezda izlashga imkon beruvchi kompyuter maʼlumotlar bazasidir. Elektron lugʻat foydalanuvchiga soʻz haqidagi barcha kerakli maʼlumotlarni iloji boricha tezroq taqdim etishni maqsad qilib qoʻyadi. Muhim urgʻu lugʻat materialining moslashuvi va elektron lugʻatning vazifalari uchun tezkor qidiruv, shuningdek, foydalanuvchiga soʻz haqida maksimal maʼlumot berishga qaratiladi, soʻzning asosiy xususiyatlari esa ikkinchi darajali narsalardan ajralib turishi kerak. Shunday qilib, elektron lugʻatlar qogʻoz lugʻatlarga qaraganda sezilarli afzalliklarga ega:

- Axborotni qayta ishlashning yuqori tezligi;
- Ekvivalentni tarjima matniga tezkor kiritish imkoniyati;
- Axborot tashuvchilarning portativligi;
- Eng yangi va maxsus leksikaning mavjudligi;
- Yangi soʻzlar bilan tezda toʻldirilishi;
- Soʻrovlar tarixini shakllantirish va boshqalar.

Elektron lugʻatlar, nafaqat kompyuter dasturi, balki koʻchma moslama ham boʻlishi mumkin. Masalan, sayyohlar orasida mashhur boʻlgan lugʻat–tarjimon yoki Quicktionary kabi skaner–tarjimon ham elektron lugʻat vazifasini bajaradi. Elektron lugʻat 2 qismdan iborat: dasturiy taʼminot qobigʻi va lugʻat bazasi. Agar qobiq deyarli oʻzgarmagan boʻlsa, unda lugʻat asoslari ulanishi, uzilishi va oʻzgarishi mumkin.

Ontologiya taʼlim sohasi boʻyicha maʼlumot almashishi kerak boʻlgan olimlar uchun umumiy soʻz boyligini belgilab beradi. Ontologik lugʻatlar – bu soʻzlar va tushunchalar oʻrtasidagi mantiqiy va semantik munosabatlar tizimini aks ettiruvchi bilimlar modelidir. Bu lugʻatlar sunʼiy intellekt tizimlarida bilimlar bazasi sifatida ishlatiladi.

WordNet lugʻatlari kompyuter lingvistikasi sohasida keng tarqalgan va ilmiy asoslangan lugʻatlardan biridir. WordNet – ingliz tilidagi leksik birliklarni ularning maʼnodoshlik va semantik munosabatlari asosida tarmoq tarzida tashkil qilgan leksik – sematik bazadir. Bu loyiha 1985 – yildan boshlab Princeton universitetida G. A. Miller rahbarligida yaratilgan va bugungi kunda koʻplab tillarda nusxalari mavjud. Tezauriyaning asosiy birligi deskriptorlar (mualliflashtirilgan atamalar) va deskriptor boʻlmaganlarga (askriptorlar) boʻlinuvchi atamalar hisoblanadi. Tezaurus lugʻatlari soʻzlar orasidagi leksik va semantik munosabatlarni aks ettiradi. Bu lugʻat faoliyatining kompyuterlashtirilishi leksikograflarning imkoniyatlarini sezilarli darajada kengaytiradi. Har bir tilning vokabulyaridan munosib oʻrin tutgan soʻzlar doimiy ravishda bir-birlari bilan paradigmatic va sintagmatic munosabatda boʻlib turadi. Shunga koʻra ularni bir necha guruhga boʻlish mumkin.

Soʻzlar aytilishi, yozilishi va maʼnosi jihatdan turlicha guruhlariga birlashishi qonuniy hisoblanadi. Shulardan biri omonimiya (shakldoshlik) hodisasidir. Aytilishi va yozilishi jihatidan bir xil, ammo maʼnosiga koʻra bir-biridan farq qiladigan soʻzlar omonimlar deyiladi. Bu soʻz grekcha homos – bir xil, onuma – nom soʻzlaridan olingan. Masalan, ot–ism, ot–hayvon, ot–uloqtirmoq; yoz–fasl, yoz–yozmoq; olma–meva, olma–olmoq feʼl va shu kabilar. Omonimiya hodisasining tayanadigan asosiy nuqtasi ham ifoda plani, tashqi tomoni sanaladi. Shu maʼnoda omonimlarning tahlili ustida jiddiy bosh qotirgan olimlarimizdan biri Sh. Rahmatullayev omonimlarni omoifodalar deb ham ataydi. Omonimlarda yana shunisi muhimki, ularni qaysi soʻz turkumiga oidligiga qarab ham ot omonimlar, feʼl omonimlar, sifat omonimlar deb guruhlariga ajratish mumkin. Sinonim–yaʼni maʼnodosh soʻzlardir. Maʼnosi bir xil shakli har xil boʻladi. Masalan, yuz–bet, aft, bashara, chehra; vatan–yurt, mamalakat, goʻsha, maskan, oʻlka va hokazolar.

Xulosa oʻrnida shuni aytish mumkinki, lugʻatlar har bir soha uchun muhim ahamiyatga ega va juda murakkab desak mubolagʻa boʻlmaydi. Oʻzbekistonda, aynan, olib borilgan tadqiqot ishlariga nazar solsak hali bu sohada ancha oqsayotganimiz va qilinishi kerak boʻlgan ishlar anchagina ekanligiga guvoh boʻlamiz.

Foydalanilgan adabiyotlar:

1. Baxtiyor Mengliyev, Shahlo Hamroyeva, Dildora Ahmedova. Kompyuter leksikografiyasi. – Toshkent: 2022
2. Abduraxmonova Nilufar. Korpus lingvistikasi (Darslik) – Toshkent:2023
3. Xudayberganova Munira. Lingvistik terminlar tezaurusini yaratishning leksikografik tadqiqi. - Tamaddun Nuri:2024
4. Poʻlatov A. Kompyuter lingvistikasi . – Toshkent: Akademnashr, 2011
5. Maxmudova Erkinoy., Abdualimova Gulhayo. Kompyuter leksikografiyasining kompyuter lingvistikasi fanidagi ahamiyati. – Образование Наука И Инновационные Идеи В Мире:2023
6. Nurmonov A., Yoʻldoshev B. Tilshunoslik va tabiiy fanlar. – Toshkent: Sharq, 2001

THEORETICAL ISSUES OF CREATING AN EDUCATIONAL CORPUS OF STUDENT SPEECH

Kuvondikova Gavhar Isomiddinovna

Teacher of the National University of Uzbekistan

E-mail: Kuvondikovag@gmail.com

Abstract: This article presents the importance of the educational corpus of students' speech and the analysis of the corpora created in this regard. Extensive work has been carried out on educational corpora of Arabic, Brazilian and Lithuanian student speech. An example of this is the development of criteria for teaching the language as a mother tongue, as a second and foreign language, and the creation of statistics on the use of academic words by students. It was also said that the project of creating an educational corpus of student speech is being implemented in Uzbekistan. In the educational corpus, students' written speech is divided into several classifications and analyzes are presented.

Keywords: Educational corpus, Arabic student corpus, Brazilian student corpus, Lithuanian corpus, written speech, classification.

Аннотация: В данной статье представлено значение учебного корпуса речи студентов и проведен анализ корпусов, созданных в связи с этим. Большая работа проведена над учебными корпусами арабской, бразильской и литовской студенческой речи. Примером этого является разработка критериев обучения языку как родному, как второму и иностранному, а также создание статистики использования академических слов учащимися. Также было сказано, что в Узбекистане реализуется проект создания учебного корпуса студенческой речи. В учебном корпусе письменная речь студентов разделена на несколько классификаций и представлены анализы.

Ключевые слова: Учебный корпус, арабский студенческий корпус, бразильский студенческий корпус, литовский корпус, письменная речь, классификация.

Introduction

A corpus is a collection of linguistic units that constitute a set of texts collected for a specific purpose. It is also a system of electronic text fragments selected on the basis of several signs that reflect the specific features and variants of a particular language, which serves as the basis for linguistic research.

Currently, a relatively young but highly developed area of corpus linguistics is the study of educational corpora of student or student speech. Such studies create a number of advantages for language learners and teachers. They also serve as a source for applied linguistics, lexicography, and other areas of linguistic research. World experience in this area is extensive and rich in diverse research directions.

Language corpus is a rapidly developing field in the world today, it is being created by corpus linguistics specialists for the purpose of conducting scientific research and teaching languages. Educational corpus is equally important in the study of the native language and a foreign language. In language teaching, corpus is very useful in showing the vastness of vocabulary, in showing an array of examples to explain the possibility of using a word through one or another grammatical construction. The constant updating of examples, which is important for language teaching, is only available in the corpus. The teacher can find new, convincing, unlimited and diverse examples here, has no difficulty in setting tasks and exercises, and can prepare tasks on the topic with new examples in a few minutes. Language corpora are an undeniable tool of modern linguistics for linguistic research and solving practical tasks. It is difficult to imagine today's theoretical and practical philology without a corpus.

Literature analysis and methods

Growing public interest in using corpora as a resource in the study of language and linguistics led to the creation of TaLC in 1994, as a result of discussions among members of ICAME (the International Computer Archive of Modern and Medieval English). The first educational corpora conference was held in Lancaster in 1994, and it was noted that the use of computer text corpora in research was well established and that they were widely used for teaching purposes, and that discussions on the use of corpus data in the preparation of teaching materials and the direct creation of corpora by students in foreign language and linguistic education were put on the agenda.

The creation of the first educational corpora was created by Sylviane Granger and her research team, the International Corpus of Learner English (ICLE). The purpose of this corpus is to classify common errors made by language learners and to develop methods for correcting them [D. Stewart: 2004].

The following characteristics of English language learners were taken into account in the creation of this educational corpus [G. Sylviane; 199]: educational environment, age, mother tongue, level of education, nature of tasks.

Language learners were selected for this corpus within the framework of the following criteria:

- 1) type of learners;
- 2) age: young adults;
- 3) level of education: advanced;
- 4) task: essays

The world experience in this area is extensive and rich in research in various directions. One of these is the ALC project, a collection of written and oral materials collected from Arabic language learners in Saudi Arabia. The goal of this project is to support the learning and teaching of Arabic as a second language and to serve as a source of information for further linguistic research.

Before the ALC corpus was created, several studies on student corpora were conducted in Saudi Arabia. These include: Pilot Arabic Learner Corpus (Abuhakema et al., 2008), Corpus of L2 (Hassan and Daud, 2011). The Pilot Arab Learner Corpus (PALC) is a corpus of Arabic texts produced by American English speakers learning Arabic, containing approximately 9,000 words. Some of the student texts were written while the students were learning Arabic in the United States. This is an intermediate level, consisting of 3,818 words. The second part is written in Arabic by those who went to study in Arab countries, and it is of high level and consists of 4741 words. Corpus of L2 Malaysian Arabic learners is a corpus of written texts by students. This corpus is mainly designed to check the incorrect use of Arabic compounds. It contains approximately 240,000 words created by 60 university students. 97% of them are Malaysians, and were collected in the first and second years of their studies in the field of Arabic. The corpus materials were developed on the topic of descriptive and comparative essays.

Based on the data from both corpora, the next draft of the student corpus, ALC, was created. This corpus was much more thorough and comprehensive than the previously created corpora. First, special attention was paid to the design of the corpus. It was planned that the ALC would consist of a series of small corpora. The design of the corpus was divided according to the following criteria:

non-native speakers of Arabic	native speakers of Arabic Arabic
Men	women
pre-university students	university students
speech materials	written materials

The information in the ALC corpus is in oral and written form, both of which consist of two stages. The first stage of the oral part was “describing a photo”, and the second stage was

“interview”. For the first stage of the written part, the topic of the story “Holiday Adventures” was chosen. Students wrote an essay in the classroom for 40 minutes. In the next stage, students wrote an essay on the topic of discussion “Should students choose education based on their interests? Or should they choose areas with a strong economic development?”. Only students had to write the essay as homework within two days, not in class or in the auditorium. Throughout the essay, students are given the opportunity to use dictionaries and grammatical information.

In Saudi Arabia, except for preschool institutions, the educational process in all other educational institutions is carried out in single-sex classes, that is, men and women do not mix. Therefore, the study paid special attention to the gender component. Only female specialists were included to work with women. Therefore, the part devoted to information about women's speech was 20%. Especially since the oral part of the corpus is laborious and time-consuming work, it was only 10% of the ALC.

The ALC corpus includes 1,585 texts (written and spoken) produced by 942 students, totaling 282,732 words. The corpus is distinguished from other corpora by its metadata. According to the metadata, the participants ranged in age from 16 to 42 and represented 67 different nationalities. The metadata of the texts is classified as follows. There are two genres in the ALC: narrative (67% of the ALC content) and discussion (%). According to the place where the texts were written: students wrote their texts in class (62% of the ALC content) and at home (31%), while all audio recordings were produced in class (7%). The average length of the texts is 178 words³³

Another major study of student speech corpora was conducted in Brazil. The researcher, Larisa Goulart Da Silva, a professor at the University of Warwick, aimed to study how Brazilian students use academic vocabulary and how corpus linguistic tools work on the basis of a pre-existing list of academic words, the AWL. The study can also be used as a source for creating a list of academic vocabulary used by Brazilian students. For this, criteria were developed for writing essays from Brazilian students studying in the UK. The essays were collected in two stages, with the first stage being descriptive essays and the second stage being argumentative essays. Larisa Goulart Da Silva initially tried to find answers to the following questions in the students' written work.

- a) What is the vocabulary profile of essays written by Brazilian students?
- b) How does it compare with the vocabulary profile of other academic British corpora?
- c) What words do Brazilian students use in the Academic Word List (AWL)?
- d) What are the differences in the use of academic words between Brazilian students and those represented in the British Academic Written English (BAWE) corpus?

The texts collected in the corpus are classified according to the following groups:

- Life Sciences,
- Social Sciences,
- Arts and Humanities,
- Physical Sciences.

We can see the level of academic word usage of students by group in the table below.

T r	Corpus	Number of running words	AWL
1	Art	875.000	9.30%
2	Science	875.000	9.10%
3	Commerce	875.000	12%
4	Law	875.000	9.40%
5	Linguistics	2.031	12.60%
6	Applied linguistics	5.137	17.60%
7	Sociology	2.084	12.44%
8	Social psychology	2.059	14.38%

9	History	2.036	14.49%
10	Devolepment	2.023	12.26%
11	Medicine	2.024	6.72%
12	Zoology	2.026	7.31%
13	Anatomy	11.356	8.60%

The researcher worked in collaboration with the Languages Without Borders (SwB) program. The program included 10,740 Brazilian students from 87 universities in the UK. These students came from 131 Brazilian universities, representing different regions of Brazil. According to information provided on the official website of the SwB, of these 10,740 students, 774 were doctoral students, 571 were full-time PhD students, 8,864 were visiting undergraduate students and 531 were postdoctoral fellows. The researcher recruited only undergraduate students for the corpus. The results of the study show that Brazilian students use academic vocabulary to the same extent as British students, but there are some specific characteristics in the use of academic vocabulary by Brazilian students. Taking into account these specific characteristics, some pedagogical recommendations for teaching EAP to Brazilian students are presented below. However, the recommendations given here mainly take into account the “Languages without Borders” program. For example, BAWE students used 43 word forms with prefixes, while Brazilian students used only one word form with prefixes, meaning that Brazilian students used less suffixes and prefixes than BAWE students. This has led to the assumption that prefixes such as “under” and “over” may not be convenient for Brazilian students’ speech, since they are not derived from Latin. However, further research is needed to confirm this hypothesis.

Studies show that there are common errors for different groups of students. For example, excessive use of conjunctions and incorrect use of lexical phrases in written text, some deviations in the use of grammatical patterns. It has been proven that the reason for the inaccuracies and hesitations in students' oral speech or the time-consuming attempts to find a solution is the fact that the speech is subject to grammatical rules, or that students cannot speak without a sequence of grammatical formulas.

Work has also begun on the creation of an educational corpus in Uzbekistan. One of these was the creation of an educational corpus of the Uzbek language based on school textbooks in 2019 under the leadership of N. Abdurakhmonova. Further research was carried out within the framework of the practical project “Creating an Educational Corpus of the Uzbek Language” under the number AM-FZ-201908172, the initial presentation of which was held on April 23, 2021, and its base is currently being enriched and improved. Today, this corpus

- 1) morphological analyzer (automatic morphological analysis);
- 2) splitting a word (form) into syllables;
- 3) providing annotated words;
- 4) displaying antonyms;
- 5) synonymizer (a program that provides synonyms for a word entered in the search);
- 6) displaying homonyms;
- 7) providing a synonym (paronym) of the word with its annotation;
- 8) providing information from the “Native Language Dictionary” related to the word entered in the search;
- 9) displaying phrases in which the word entered in the search is present;
- 10) It has the ability to display a ranking series based on its various characteristics.

Practical significance

Experience in world linguistics shows that various studies have been carried out to enrich the educational corpus, enrich its base with more useful and necessary information for the student. The most popular of these studies is the study of student speech and its inclusion in the corpus. Using the positive results and experiences achieved in the educational corpora of student speech

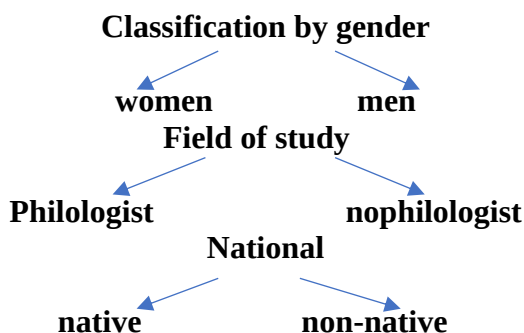
studied in linguistics, we have begun research on the creation of corpora of student speech in the national corpus of the Uzbek language and the formation of methodological developments in this regard. Uzbek linguistics has not yet paid attention to studies in this regard. The project is being implemented at the Department of Computational Linguistics and Applied Linguistics of the National University of Uzbekistan. The main objectives of the research are as follows:

I. Aims to create methods for correcting common errors in written speech by classifying them

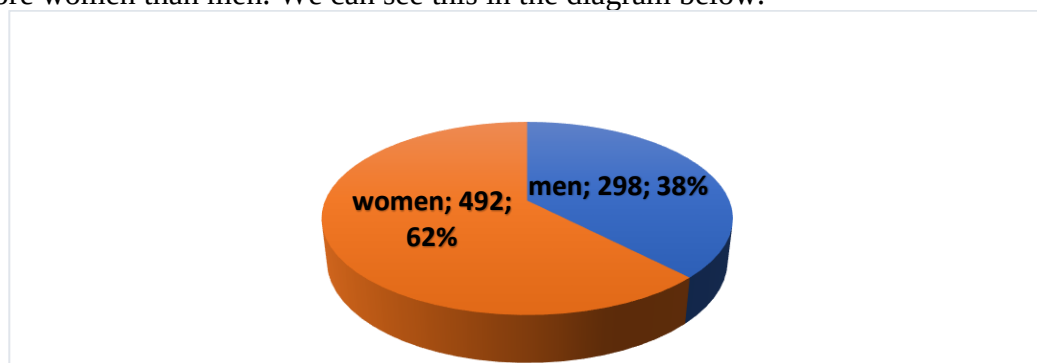
II. To be useful in the study and teaching of the Uzbek language as a first and second language.

III. To serve as a source for further research in various fields of linguistics.

The initial work was on the form of the study. In order to clearly imagine what processes are taking place in the language, it was considered appropriate to use written speech materials in the corpus. Essays from undergraduate students were selected as the research material. Students of undergraduate students from various faculties and areas of the National University were included. The essay procedure was carried out in two stages. The first stage was organized during the 1st, 3rd, 5th, and 7th semesters of the students. The topics selected for the first stage of the study were “Does a good university diploma guarantee a good job” and “Today, many people prefer distance learning to face-to-face education.” It was decided that students would be free to choose one of the two topics. The essay should be written in the classroom during the 40-minute lecture. The topic of the second stage of writing was selected under the title of “People living in big cities face many problems. Should the government encourage them to move to other cities?” in the discussion content. We planned to carry out the research in the next stage in the 2nd, 4th, 6th and 8th semesters. In order to achieve the intended goal of the research, it is necessary to classify into groups. The world experiences in dividing into groups were classified as follows based on the British Academic Written English (BAWE) and the Arabic Learner Corpus (ALC):



In the first stage of our study, 790 students from 15 areas of 10 faculties of the National University participated. We begin to analyze the collected written work samples based on the 1st classification. According to the results, among the students who participated in the first stage, there were more women than men. We can see this in the diagram below.



If we look at the next results in the first classification, we see the statistics of the choice of the topics given to students: “Does a good university degree guarantee a good job” and “Today, many people prefer distance learning to face-to-face education.” The need for this statistics will serve as a basis for the next stages of the study. For example, it plays an important role in identifying the most frequently used words in students' written speech. In addition, it is also important in the processes of grammatical and syntactic analysis of written materials. According to the results:

The topic “Does a good university degree guarantee a good job?” was chosen by 561 students, and the topic “Nowadays, many people prefer distance learning to face-to-face education” was chosen by 229 students.

Conclusion.

The work on creating Uzbek language corpora in our country is not satisfactory. The scientific research work being carried out in this area is at the stage of developing the theoretical foundations of Uzbek corpus linguistics, creating one or two corpus sheets as examples. The goal is to develop Uzbek corpus linguistics, which is considered a direction of applied linguistics in Uzbekistan, to study it scientifically and theoretically, to establish a center and laboratory for computer and corpus linguistics. The large-scale nature of the corpus guarantees the originality of the information, and ensures the full demonstration of all aspects of linguistic phenomena. Various types of information are placed in the language corpus in their natural contextual form, which becomes the basis for its comprehensive and objective study. In order to enrich the information in the corpus and increase its usefulness, we aimed to create an educational corpus of student speech.

References:

1. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL). 2019;6(1-2019):131-7.
2. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. In Proceedings of the International Conference on Language Technologies for All (LT4All) 2019.
3. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL), 6(1-2019), 131-7. 7. Абдурахмонова, Н., & Абдувахобов, Г. (2021). O‘quv lug‘atini tuzishning nazariy metodologik asoslari. Международный журнал искусство слова, 4(6)
4. Abdullah Alfaifi, Eric Atw, Hedeya Ibrahim. Arabic Learner Corpus (ALC) v2. A New Written and Spoken Corpus of Arabic Learner 2014 Saudia Arabia
5. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies, 10(8), 59-66.
6. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In 2022 7th International Conference on Computer Science and Engineering (UBMK) (pp. 73-75). IEEE.
7. Larisa Goulart Da Silva Academic vocabulary: a corpus linguistics study on how Brazilian students write academic English. 2016-a. Brazilia
8. Granger, S. (2003a). Error-tagged learner corpora and CALL: A promising synergy. CALICO Journal, 20(3): 465-480.
9. Hammarberg, B. (2010). Introduction to the ASU corpus, a longitudinal oral and written text corpus of adult learners' Swedish with a corresponding part from native Swedes. Stockholm University: Department of Linguistics.
10. Thompson, P. (2005). Spoken language corpora. In M. Wynne (Ed.), Developing Linguistic Corpora: a Guide to Good Practice (pp. 59-70). Oxford: Oxbow Books.

O'ZBEK TILI MORFOLOGIYASINING ONTOLOGIK MODEL

Sobirova Madina

O'zbekiston Milliy Universiteti magistrantlari

Madadova Afag

O'zbekiston Milliy Universiteti magistrantlari

Annotatsiya: O'zbek tili morfologiyasining ontologik modelini ishlab chiqish masalasi bugungi kunda NLP va mashina tarjimasida sohalarida alohida ahamiyatga ega mavzudir. Shu bois, tadqiqot davomida mazkur mavzu yoritiladi. Unda so'z turkumlari, ularning morfologik xususiyatlari va o'zaro munosabatlari konseptual asosda tizimlashtiriladi. Shuningdek, ot so'z turkumi uchun ontologik model namunasi ishlab chiqilib, ularning semantik va grammatik belgilari aniqlanadi. Tadqiqot natijalari o'zbek tilining kompyuter lingvistikasi, NLP va sun'iy intellekt tizimlarida qo'llash imkoniyatlarini kengaytirishga xizmat qiladi.

Kalit so'zlar: ontologik model, morfologiya, akronimlar, leksikografiya, NLP, kompyuter lingvistikasi.

Annotation: The development of an ontological model of Uzbek language morphology is currently a particularly important topic in the fields of NLP and machine translation. Therefore, this study addresses this issue. It systematizes parts of speech, their morphological features, and their interrelations based on a conceptual framework. Additionally, an example of an ontological model is developed for the noun part of speech, with its semantic and grammatical attributes identified. The results of this research contribute to expanding the application opportunities of the Uzbek language in computer linguistics, NLP, and artificial intelligence systems.

Keywords: ontological model, morphology, acronyms, lexicography, NLP, computational linguistics.

Аннотация: Разработка онтологической модели морфологии узбекского языка в настоящее время является особенно актуальной темой в области NLP и машинного перевода. В связи с этим в данном исследовании рассматривается данный вопрос. В работе систематизируются части речи, их морфологические особенности и взаимосвязи на концептуальной основе. Кроме того, разрабатывается пример онтологической модели для именной части речи с определением её семантических и грамматических характеристик. Результаты исследования способствуют расширению возможностей применения узбекского языка в компьютерной лингвистике, NLP и системах искусственного интеллекта.

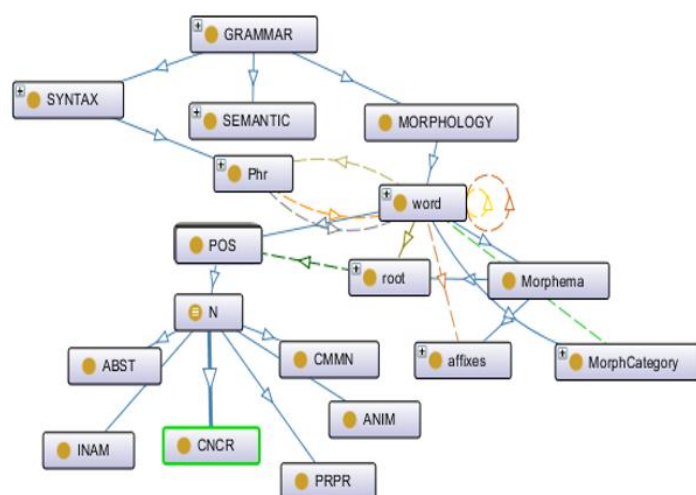
Ключевые слова: онтологическая модель, морфология, акронимы, лексикография, NLP, компьютерная лингвистика.

Hozirgi texnika asri tilshunosligida tabiiy tillarning kompyuter modellari va ularning ontologik asoslarini yaratish dolzarb masalalardan biri hisoblanadi. Xususan, o'zbek tilining morfologik tizimini ontologik model asosida tasvirlash — til birliklarini tartibga solish, ularning o'zaro munosabatlarini aniqlash hamda kompyuter lingvistikasi va sun'iy intellekt tizimlarida samarali foydalanish imkonini beradi. Ushbu maqolada o'zbek tili morfologiyasining asosiy so'z turkumlaridan biri ot so'z turkumi va uning grammatik xususiyatlarini ontologik model orqali tasvirlash masalasi ko'rib chiqiladi. Dastavval, ontologiyaga professorlar tomonidan berilgan ta'riflar orqali bu termin ma'nosini chuqurroq bayon qilinadi. Ontologiya — bu konseptualizatsiyaning spetsifikatsiyasi[Gruber, T. R., 1993]. Demak, ilk qarashlarda ontologiya ma'lum bir sohadagi tushunchalar va ularning o'zaro munosabatlarini rasmiy tarzda ifodalovchi vosita sifatida qaraladi. Ontologiya — bu mavjudotlar va ularning mavjud bo'lishi mumkin

bo'lgan shakllarini o'rganadigan fan[Guarino, N., 1998, 98]. Guarino ontologiyani tushunchalar va ularning o'zaro aloqalarini aniqlash orqali bilimlarni modellashtirish vositasi sifatida ta'riflaydi. Ontologiya — bu tushunchalar to'plami va ularning o'zaro aloqalarini ifodalovchi yaqqol spetsifikatsiyadir[Noy, N. F., & McGuinness, D. L., 2001]. Bu ta'rifda ontologiya bilimlarni almashish va qayta foydalanish uchun umumiy tushunchalar va munosabatlar to'plami sifatida ko'riladi. Ontologiyalar — bu tushunchalar va ularning o'zaro aloqalarini rasmiy tarzda ifodalovchi vositalar bo'lib, ular avtomatlashtirilgan tizimlarda bilimlarni boshqarish va almashish imkonini beradi[Staab, S., & Studer, R. 2009]. Aynan shu munosabat bilan ontologiyalar tabiiy tilni qayta ishlash sohasida alohida ahamiyatga ega. Shu jumladan, o'zbek tili grammatikasining ontologiyasi tuzilishi ham bu sohadagi birqancha jumboqlarga yechim bo'ladi. Bu ishning amaliy jihatini morfologiya misolida ko'ramiz. Umuman olganda, har qanday tabiiy til morfologiyasida, jumladan o'zbek tilida ham, eng asosiy so'z turkumi bu — ot hisoblanadi. Shu sababli ot so'z turkumining ontologik modelini birinchi navbatda yaratish maqsadga muvofiq bo'ladi. O'zbek tilidagi otni sinf sifatida ontologik modelda tasvirlash uchun quyidagi qisqacha tushuntirishdan foydalanamiz. Ontologik modellarda ingliz tilidagi akronimlar – ingliz tili grammatik kategoriya nomlarining qisqartmalari mavjud bo'lib, boshqa barcha tillar o'z grammatikasidan kelib chiqib o'zlarinikiga mostlashtiradilar. Quyidagi jadval orqali ularga misollar keltirilgan.

Akronimlar	To'liq shakli(inglizcha)	Grammatik kategoriya nomlari
POS	Part of speech	So'z turkumlari
N	Noun	Ot
PRPR	Proper	Atoqli
CMMN	Common	Turdosh
ABST	Abstract	Mavhum
CNCR	Concrete	Aniq
ANIM	Animate	? Shaxs oti, narsa oti
INANIM	Inanimate	? O'rin- joy oti, narsa oti

Jadvalda ko'rinib turganidek, ingliz tili va bizning ona tilimiz o'rtasida ayrim farqlar mavjud. Xususan, jonli va jonsiz otlar degan tushuncha o'zbek tili grammatikasida yo'q. O'rniga yanada qulayroq shaxs, narsa va joy oti turlariga bo'lingan. Jahon standartiga moslashtirish uchun bu uch turni ikkiga bo'lishni tavsiya qilgan bo'lardim. Biroq buning uchun o'zbek tili leksikografiyasida ham jonli va jonsiz otlar lug'ati tuzilishi kerak. Bundan tashqari, otlar ko'plik, egalik, kelishik shakllarida tuslana oladi. Shuning uchun ularni tavsiflovchi munosabatlar va funksiyalar aniqlanadi va ular yo'naltirilgan chiziqlar bilan ifodalanadi. Mazkur qabul qilingan belgilar va kelishuvlardan foydalanib, o'zbek tilidagi otning ontologik modelini tuzamiz. Bunda bizga protégé dasturi yaqindan yordam beradi.



1-rasm. O'zbek tilidagi otning ontologik modeli.

Mazkur tadqiqot natijasida o‘zbek tili morfologiyasining ontologik modelini ishlab chiqish bo‘yicha konseptual asoslar shakllantirildi. So‘z turkumlari, ularning morfologik xususiyatlari va o‘zaro munosabatlari tizimli tarzda tahlil qilinib, ot so‘z turkumi uchun ontologik model namunasi ishlab chiqildi. Modelda otlarning semantik va grammatik belgilari aniq belgilandi. Tadqiqot natijalari o‘zbek tilining kompyuter va korpus lingvistikasi, tabiiy tilni qayta ishlash (NLP) va sun‘iy intellekt tizimlarida samarali qo‘llash imkoniyatlarini kengaytirishga xizmat qilishi, shuningdek, mashina tarjimai va lingvistik tahlil jarayonlarini yanada takomillashtirishga zamin yaratishi mumkin.

Foydalanilgan adabiyotlar.

1. Gruber, T. R. (1993). *A translation approach to portable ontology specifications*. Knowledge Acquisition, 5(2), 199–220.
2. Guarino, N. (1998). *Formal ontology and information systems*. In *Proceedings of FOIS* (Vol. 98, pp. 3–15).
3. Noy, N. F., & McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-5.
4. Staab, S., & Studer, R. (2009). *Handbook on Ontologies*. Springer Science & Business Media.
5. Abduraxmonova, N. Z. (2023). *Korpus lingvistikasi*. Toshkent: Globeedit. 346 b
6. Abduraxmonova, N. Z., Qodirova, Z. (2023) *Kompyuter leksikografiyasida ontologiya(monografiya)*. Toshkent: O‘zbekiston Milliy universiteti. 104b
7. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, 11(19), 9117.
8. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).

DIAXRONIK KORPUSLAR SHAKLLANISH EVOLYUTSIYASI

Askarov Ilhom

Mirzo Ulug‘bek nomidagi

O‘zbekiston Milliy universiteti 2-bosqich magistranti

E-mail: asqarovilhomjon21@gmail.com

Annotatsiya: Ushbu maqolada diaxronik korpuslar va ularning lingvistik tadqiqotlardagi o‘rni tahlil qilinadi. Korpuslarning tarixiy rivojlanishi 1960-yillardan boshlab zamonaviy raqamli texnologiyalar bilan bog‘liq holda bosqichma-bosqich ko‘rib chiqiladi. Maqolada ingliz, fransuz, nemis va ispan tillaridagi mashhur diaxronik korpuslar misolida til o‘zgarishlarining statistik va kontekstual tahlili yoritilgan. Shuningdek, korpuslar bilan ishlashda foydalaniladigan dasturiy vositalar va ularning imkoniyatlari haqida ma’lumot beriladi. Diaxronik korpuslar tilshunoslikda tarixiy tahlil va til taraqqiyotini chuqur o‘rganish imkonini beradi.

Kalit so‘zlar: diaxronik korpus, korpus lingvistikasi, korpus evalyutsiyasi, lingvistik instrumentary.

Annotation: This article analyzes diachronic corpora and their role in linguistic research. The historical development of these corpora is examined step by step, beginning from the 1960s in connection with modern digital technologies. The article highlights statistical and contextual analysis of language change using well-known diachronic corpora in English, French, German, and Spanish. It also provides information about the software tools used in working with corpora and their capabilities. Diachronic corpora enable an in-depth study of historical analysis and language development in linguistics.

Keywords: diachronic corpus, corpus linguistics, corpus evolution, linguistic tools.

Аннотация: В статье анализируются диахронические корпуса и их роль в лингвистических исследованиях. Историческое развитие этих корпусов рассматривается поэтапно, начиная с 1960-х годов в контексте современных цифровых технологий. На примере известных диахронических корпусов на английском, французском, немецком и испанском языках освещается статистический и контекстуальный анализ языковых изменений. Также представлена информация о программных инструментах, используемых при работе с корпусами, и их возможностях. Диахронические корпуса позволяют глубоко изучать исторический анализ и развитие языка в лингвистике.

Ключевые слова: диахронический корпус, корпусная лингвистика, эволюция корпусов, лингвистические инструменты.

Diaxronik korpus — bu tildagi vaqtinchalik o‘zgarishlarni, ya’ni tilning tarixi davomida qanday o‘zgarishini o‘rganish uchun yaratilgan matnlar to‘plamidir. “Diaxronik” so‘zi yunoncha “dia” (orqali) va “chronos” (vaqt) so‘zlaridan kelib chiqqan bo‘lib, tilshunoslikda til hodisalarini vaqt davomida, tarixiy taraqqiyot jarayonida o‘rganishni anglatadi. 1990-yillarda kompyuter texnologiyalarining rivojlanishi tilshunoslikda avtomatik tahlil qilish imkoniyatlarini yaratdi.

Diaxronik korpuslarning evolyutsiyasi bir necha bosqichlardan o‘tgan:

Erta davr: Dastlabki diaxronik korpuslar odatda yozma adabiyotlar hujjatlarni o‘z ichiga olgan. Bu matnlar ko‘pincha butunlay qo‘lda qisman kompyuter yordamida tahlil qilingan.

Kompyuterlashtirish bosqichi: 1990-yillarda, kompyuter texnologiyalarining rivojlanishi bilan, tilshunoslar va lingvistlar diaxronik korpuslarni yaratish uchun kompyuter dasturlaridan foydalanishni boshladilar. Bu korpuslar hozirda turli tillar bo‘yicha tez va samarali tahlil qilishni ta’minlaydi.

Kengaytirilgan korpuslar: 2000-yillardan boshlab, diaxronik korpuslar yanada kengayib, ko'proq tillar va ma'lumot manbalarini o'z ichiga olgan. Shu bilan birga, korpuslar tez-tez yangilanib turadi.

Dastlabki bosqichda, foydalanuvchilarga davrlar bo'yicha kichik namunalar bilan ishlash va matnlarga tegishli meta-ma'lumotlarni diqqat bilan o'rganish tavsiya etiladi. Korpus bilan ishlashda AntConc¹⁴¹ Sketch Engine¹⁴² kabi maxsus dasturiy vositalardan foydalanish mumkin. Shuningdek, natijalarni XLS formatida eksport qilish imkoniyati mavjud bo'lib, bu tahlil jarayonini yanada qulaylashtiradi. Boshqa tillarda ham ilk diaxronik korpuslar yaratilib, til tarixi va o'zgarishlarini o'rganishda muhim manba bo'lib xizmat qilgan. Fransuz tilida yaratilgan dastlabki korpuslardan biri — **Frantext**¹⁴³, 1970-yillarda tashkil topgan bo'lib, XIV–XX asrlarga oid fransuz adabiyoti matnlarini o'z ichiga oladi. Shuningdek, **Base de Français Médiéval**¹⁴⁴ nomli korpus o'rta asr fransuz tiliga ixtisoslashgan bo'lib, tarixiy tilshunoslik uchun muhim resurs hisoblanadi. Nemis tilida esa **Bonner Frühneuhochdeutsch Korpus**¹⁴⁵ 1980-yillarda yaratilgan bo'lib, 1350–1700-yillar oralig'idagi nemis tilini qamrab oladi. Bundan tashqari, **Referenzkorpus Altdeutsch**¹⁴⁶ qadimgi yuqori nemis tili matnlarini jamlagan bo'lib, nemis tilining tarixiy bosqichlarini tadqiq qilishda muhim ahamiyat kasb etadi.

Dastlabki diaxronik korpuslar bir qator metodologik cheklovlar bilan ajralib turadi. Texnik jihatdan, bu korpuslar odatda nisbatan kichik hajmda bo'lgan (1–5 million so'z atrofida) va o'sha davrda matnlarni skanerlash hamda raqamlashtirish texnologiyalari yetarlicha rivojlanmaganligi sababli, ma'lumotlarni to'plash va tayyorlash jarayoni murakkab kechgan. Annotatsiya borasida ham muammolar mavjud bo'lgan — avtomatik teglash algoritmlari hozirgi darajada bo'lmaganligi sababli, ko'plab matnlar qo'lda annotatsiya qilingan, bu esa mehnat talab qiluvchi va vaqt oluvchi jarayon hisoblangan. Diaxronik korpuslarning evolyutsiyasi muayyan bosqichma-bosqich taraqqiyot yo'lini bosib o'tgan bo'lib, ularning shakllanishi va takomillashuvi turli davrlardagi ilmiy-texnik yutuqlar, lingvistik yondashuvlar va raqamli texnologiyalarning rivojlanishiga bevosita bog'liq holda quyidagi bosqichlarga bo'lish mumkin:

Dastlabki bosqich (1960–1980-yillar) kompyuter texnologiyalarining rivojlanishi matnlarni elektron shaklda saqlash imkonini yaratdi. Natijada ingliz va nemis tillarida kichik korpuslar tuzilib, tilshunoslikda yangi metodologiyalar paydo bo'ldi hamda til materiallarini raqamlashtirish jarayoni boshlandi.

Kengaytirilgan bosqich (1990–2000-yillar) davomida ko'proq tillar, jumladan, fransuz, ispan va rus tillari uchun diaxronik korpuslar ishlab chiqildi, shu bilan birga, Amerika ingliz tilining 1810-2009-yillardagi matnlarini o'z ichiga olgan 400 million so'zlik The Corpus of Historical American English (COHA)¹⁴⁷ va rus tilining turli davrlaridagi matnlarini qamrab olgan Russian National Corpus (RNC)¹⁴⁸ kabi yirik diaxronik korpuslar yaratildi.

Zamonaviy bosqich (2010-yillardan hozirgi kungacha) kompyuter lingvistikasi va korpus lingvistikasi sohasida katta yutuqlarga erishilgan davr hisoblanadi. Ushbu bosqichda 1 milliard so'zdan ortiq bo'lgan katta hajmdagi korpuslar yaratildi. Masalan, **Google Books Ngram Corpus**¹⁴⁹ 1800-2000-yillarga oid 8 million kitob asosida tuzilgan bo'lib, tilning tarixiy rivojlanishini o'rganishda muhim manba hisoblanadi. Shuningdek, **The Corpus of Middle**

¹⁴¹ Anthony, L. AntConc (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Available from: <https://www.laurenceanthony.net/software/antconc/>

¹⁴² Kilgarriff A and others. The Sketch Engine: Ten Years On //Lexicography, 2014 (1/1). – P. 7-36.

¹⁴³ Grefenstette G. (1994). Frantext: A French Text Corpus for Linguistic Research. In Proceedings of the 15th International Conference on Computational Linguistics (COLING 1994). – Kyoto, Japan.

¹⁴⁴ <http://bfm.ens-lyon.fr/>

¹⁴⁵ <https://www.germanistik-im-netz.de/informieren/detail/ressource/das-bonner-fruehneuhochdeutschkorpus>

¹⁴⁶ Mittmann Roland and Plate Ralf. "Das Referenzkorpus Altdeutsch" als Lesekorpus. Grammatisch annotierte und mit Wörterbüchern verknüpfte Texte für Lehre und Selbststudium" Das Mittelalter, vol. 24, no. 1, 2019. – P. 173-187. <https://doi.org/10.1515/mial-2019-0012>

¹⁴⁷ Davies Mark. (2012). Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. Corpora, 7(2), 121–157. <https://doi.org/10.3366/cor.2012.0024>

¹⁴⁸ <https://ruscorpora.ru/en/page/corpora-datasets>

¹⁴⁹ <https://www.wired.com/2015/10/pitfalls-of-studying-language-with-google-ngram>

English Prose and Verse¹⁵⁰ oʻrta ingliz tilining barcha janrlarini qamrab olib, oʻrta asr ingliz adabiyotini tahlil qilishda keng qoʻllaniladi.

Diachronik korpuslar tilshunoslarga tarixiy kontekstda tilning qanday rivojlanganini oʻrganishda yordam beradi. Shu orqali, tilning yirik tarixiy davrlaridagi oʻzgarishlarni, masalan, Rennessans, Oltin asrlar industrial inqilob davridagi oʻzgarishlarni koʻrib chiqish mumkin

Baʼzi mashhur diachronik korpuslar quyidagilardan iborat:

- **Helsinki Corpus** (Ingliz tili): Ingliz tilining tarixiy oʻzgarishlarini oʻrganish uchun keng qamrovli diachronik korpus.
- **Deutsches Textarchiv**¹⁵¹ (Nemis tili): Nemis tilidagi matnlarning diachronik toʻplami.
- **Corpus del Español**¹⁵² (Ispan tili): Ispan tilining tarixiy rivojlanishini oʻrganish uchun yaratilgan korpus.
- **Lancashire Corpus**¹⁵³ (Ingliz tili): Ingliz tilidagi xalq tili va leksik oʻzgarishlarni oʻrganish uchun foydalidir.

Helsinki Korpusi — ingliz tilidagi matnlar toʻplami boʻlib, lingvistik, kompyuter lingvistikasi va NLP sohalarida qoʻllaniladi. 1991-yilda Finlandiyaning Helsinki universitetida Matti Rissanen va jamoadoshi olimlar tomonidan yaratilgan. Korpus 700–1710-yillar oraligʻidagi matnlarni oʻz ichiga oladi va morfologik, sintaktik hamda semantik tahlil uchun moʻljallangan. Unda 450 manbadan olingan taxminan 1,5 million soʻz mavjud.

Korpus 3 asosiy davrga boʻlingan:

Davr nomi	Yillar	Soʻzlar	Asosiy xususiyatlar
Old English	700-1150	413,250	Anglosakson davri matnlari
Middle English	1150-1500	608,570	Norman istilosidan keyingi davr
Early Modern English	1500-1710	551,000	Uygʻonish davri va keyingi davr

Deutsches Textarchiv (DTA) — XVIII–XX asrlardagi nemis tilidagi tarixiy matnlarni oʻz ichiga olgan korpus boʻlib, nemis tili va adabiyotining rivojlanishini oʻrganishga moʻljallangan. Korpusda adabiy, ilmiy, tarixiy va jurnalistik matnlar jamlangan boʻlib, u millionlab soʻzlarni qamrab oladi.

Corpus del Español (Ispan tili korpusi) — bu ispan tilining tarixiy, madaniy va geografik xilma-xilligini oʻrganish uchun yaratilgan katta hajmdagi matnlar toʻplami. Korpus 1200–2000-yillarda yozilgan matnlarni oʻz ichiga oladi va Ispaniya hamda Lotin Amerikasining turli hududlaridan olingan adabiy, ilmiy va publitsistik janrdagi matnlarni qamrab oladi.

Adabiy asarlar: korpusda ispan tilidagi eng mashhur adabiyotlar va klassik asarlar mavjud, jumladan, Migel de Cervantesning “Don Kixot”i, Lorca, Quevedo, va boshqa mashhur yozuvchilar asarlari.

Jurnalistik matnlar: gazeta va jurnallardan olingan maqolalar, intervyular va boshqa ommaviy axborot vositalari materiallari.

Ilmiy hujjatlar: turli ilmiy va akademik maqolalar, tadqiqotlar va monografiyalar.

- **Frantext** — bu XVI asrdan to zamonaviy davrgacha boʻlgan fransuz tilidagi matnlarni oʻz ichiga olgan yirik korpusdir. U adabiy, ilmiy va jurnalistik asarlar asosida tuzilgan boʻlib, fransuz tili tarixi, leksikasi, sintaksisi va stilistikasi boʻyicha tadqiqotlar uchun muhim manba hisoblanadi.

Frantext 2 milliondan ortiq soʻz va 5000 dan ziyod fransuz tilidagi asarni oʻz ichiga olgan katta hajmdagi korpusdir. U adabiy, ilmiy va jurnalistik matnlar asosida tuzilgan boʻlib, matnlar tematik va tarixiy davrlarga qarab tashkil etilgan. U Molière, Voltaire, Flaubert, Zola kabi mashhur yozuvchilarning asarlarini chuqur tahlil qilish imkonini beradi.

¹⁵⁰ <https://www.lib.umich.edu/Corpus of Middle English Prose and Verse>

¹⁵¹ <https://deutsches-textarchiv.de/>

¹⁵² https://americandialect.org/corpus_del_espaol/

¹⁵³ Lancashire Corpus. (n.d.). Trinity Lancaster Corpus. Retrieved from <https://cass.lancs.ac.uk/trinity-lancaster-corpus/>

Xulosa qilib aytganda, diaxronik korpuslar tilshunoslikda tilning tarixiy rivojlanishini tahlil qilish uchun muhim manba hisoblanadi. Ular turli tarixiy davrlardagi til o'zgarishlarini aniqlashga yordam beradi. Korpuslar texnologik taraqqiyot bilan birga shakllanib, keng ko'lamli raqamli tahlillarni amalga oshirish imkonini yaratdi. Maxsus dasturiy vositalar yordamida korpuslar bilan samarali ishlash mumkin. Bu korpuslar tilshunoslarga leksik, grammatik va stilistik o'zgarishlarni chuqur o'rganish imkoniyatini beradi.

Foydalanilgan adabiyotlar:

1. Gatiatullin, A., Suleymanov, D., Prokopyev, N., & Abdurakhmonova, N. (2022, November). "Turkic Morpheme": From the Portal to the Linguistic Platform. In World Conference Intelligent System for Industrial Automation (pp. 181-189)
2. Anthony, L. AntConc (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Available from: <https://www.laurenceanthony.net/software/antconc/>
3. Kilgariff A and others. The Sketch Engine: Ten Years On //Lexicography, 2014 (1/1). – P. 7-36.
4. Grefenstette G. (1994). Frantext: A French Text Corpus for Linguistic Research. In Proceedings of the 15th International Conference on Computational Linguistics (COLING 1994). – Kyoto, Japan.
5. <http://bfm.ens-lyon.fr/>
6. Mittmann Roland and Plate Ralf. "Das Referenzkorpus Altdeutsch' als Lesekorpus. Grammatisch annotierte und mit Wörterbüchern verknüpfte Texte für Lehre und Selbststudium" Das Mittelalter, vol. 24, no. 1, 2019. – P. 173-187. <https://doi.org/10.1515/mial-2019-0012>
7. Davies Mark. (2012). Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. Corpora, 7(2), 121–157. <https://doi.org/10.3366/cor.2012.0024>
8. <https://ruscorpora.ru/en/page/corpora-datasets>
9. Lancashire Corpus. (n.d.). Trinity Lancaster Corpus. Retrieved from <https://cass.lancs.ac.uk/trinity-lancaster-corpus/>
10. <https://www.wired.com/2015/10/pitfalls-of-studying-language-with-google-ngram>
11. <https://www.lib.umich.edu/Corpus of Middle English Prose and Verse>
12. <https://deutsches-textarchiv.de/>
13. https://americandialect.org/corpus_del_espaol/
14. <https://www.germanistik-im-netz.de/informieren/detail/ressource/das-bonner>

IKKI TILDAGI FRAZEOLOGIK BIRIKMALARNING PARALLEL KORPUSDA BERILISHI

Abduraxmonov Umbar

Mirzo Ulugʻbek nomidagi

Oʻzbekiston Milliy universiteti 2-bosqich magistri

E-mail: umbarabdurahmonov@gmail.com

Annotatsiya: Ushbu maqolada parallel korpuslarning frazeologik birliklarni tadqiq etishdagi roli yoritiladi. Ingliz va oʻzbek tillaridagi frazeologik iboralar semantik, sintaktik va madaniy jihatdan farqlanishi qayd etiladi. Parallel korpuslardan foydalanish frazeologik iboralarning tarjima variantlarini aniqlash, ularning kontekstdagi qoʻllanilishini tahlil qilish hamda tarjima strategiyalarini oʻrganish imkonini beradi. .

Kalit soʻzlar: parallelkorpus, frazeologik birikma, fraza, matn segmentatsiyasi.

Annotation: This article explores the role of parallel corpora in the study of phraseological units. It highlights the semantic, syntactic, and cultural differences between English and Uzbek idioms. The use of parallel corpora enables the identification of translation variants, analysis of contextual usage, and study of translation strategies.

Keywords: parallel corpus, phraseological unit, phrase, text segmentation.

Аннотация: В данной статье рассматривается роль параллельных корпусов в изучении фразеологических единиц. Отмечаются семантические, синтаксические и культурные различия между идиомами английского и узбекского языков. Использование параллельных корпусов позволяет выявлять варианты перевода, анализировать контекстуальное употребление и изучать переводческие стратегии.

Ключевые слова: параллельный корпус, фразеологическая единица, фраза, сегментация текста

Parallel korpuslar – ikki yoki undan ortiq tillarda bir-biriga moslashtirilgan matnlar toʻplami – soʻnggi yillarda tilshunoslar, tarjimonlar va informatika mutaxassisleri tomonidan keng qoʻllanilmoqda.

Frazeologik tadqiqotlarda parallel korpuslardan foydalanish quyidagi imkoniyatlarni beradi:

- frazeologik iboralarning aniq tarjima variantlarini aniqlash;
- kontekstda qoʻllanilishining solishtirilishi;
- tarjima strategiyalarini oʻrganish;
- til oʻrgatuvchi va tarjimonlar uchun amaliy qoʻllanma tayyorlash.

“Tilshunoslar nutqining grammatik talqinini isbotlash korpus tilshunosligi davri uchun endi yangilik emas edi. Odatdagi holatlar, korpus tadqiqotlaridagi oʻn, yuz va minglab misollarning misollari tadqiqotchiga xulosa chiqarishga yordam beradi; noodatiy vaziyatlar ham yangi nazariy qarashlarni yuzaga keltiradi. Boshqacha aytganda, nutqdagi yangi kashfiyotlar, yangi birikmalar va tuzilmalar tilshunoslikda yangi farazlarga imkon beradi”¹⁵⁴.

Hozirgi kunda korpus lingvistikasi va parallel korpuslar tilshunoslik, mashina tarjimasi va til oʻrganish sohalarida muhim ahamiyatga ega. Oʻzbekistonda ham ushbu sohada bir qator tadqiqotlar amalga oshirilmoqda. Misol uchun, Oʻzbekistonda parallel korpuslarni yaratish va ularni tahlil qilish boʻyicha bir qator ilmiy ishlar olib borilmoqda. Botir Elov va Maʼrufjon Amirqulov tomonidan oʻtkazilgan tadqiqotda oʻzbek-ingliz tillarining teglangan parallel korpusini

¹⁵⁴ Karimov Rustam, Mengliyev Baxtiyor. The Role of the Parallel Corpus in Linguistics, the Importance and the Possibilities of Interpretation. – International Journal of Engineering and Advanced Technology (IJEAT). 2019. https://tsuull.uz/sites/default/files/the_role_of_the_parallel_corpus_in_linguistics.pdf

yaratish bosqichlari va ularga qo'yiladigan lingvistik va ekstralingvistik talablar ko'rib chiqilgan¹⁵⁵. Shuningdek, Abduraxmonova Nilufar va Axmedova Anorxon Nasivali qizi tomonidan parallel va qiyosiy korpuslarning tarjimashunoslik va yozma tarjimoni o'qitishdagi o'rni va amaliy ahamiyati tahlil qilingan. Bu kabi tadqiqotlar parallel korpuslarning yaratilishi va ularning lingvistik tahlilidagi muammolarni hal etishda muhim ahamiyatga ega.

Frazeologik birliklar – bu tilning to'g'ri ishlatish shartlari mavjud bo'lgan, o'zgarishga unchalik moyil bo'lmagan so'z birikmalaridir. Ular tilning barqaror strukturalarini tashkil etadi. Ingliz va o'zbek tillarida frazeologik birliklar har xil semantik va sintaktik strukturaga ega, shuningdek, ularning ko'pida madaniy farqlar mavjud. Bu frazeologik birliklar parallel korpusda qanday aks etadi, ularni tarjimashunoslikda qanday ishlatish mumkinligi alohida tadqiqotlarni talab qiladi.

Inglizcha frazeologizm	O'zbek tilidagi ekvivalenti	Lug'aviy tarjima	Izoh
Break a leg	Omad tilamoq	Oyoqni sindirish	Musobaqa yoki o'yindan oldin omad tilash
The ball is in your court	Endi navbat sendan	To'p sening maydoningda	Mas'uliyat yoki navbat kimdadir ekanligini bildiradi
Hit the sack	Uyquga ketmoq	Xaltani urmoq	Yotishga ketmoq, dam olmoq
Let the cat out of the bag	Sirni ochmoq	Mushukni sumkadan chiqarmoq	Sirni bilmasdan aytib qo'yish
Once in a blue moon	Onda-sonda	Ko'k oy chiqqanda	Juda noyob yoki kamdan-kam hollarda sodir bo'ladigan voqea
A piece of cake	Xamirdan qil sug'urganday	Pirog bo'lagi	Juda oson yoki soddalashtirilgan vazifa
Spill the beans	Sirni fosh qilmoq	Loviya to'kmoq	Yashirin narsani aytib qo'yish
Bite the bullet	Tishini tishiga qo'ymoq	O'qni tishlamoq	Og'riqli yoki yoqimsiz vaziyatni qabul qilish
Cost an arm and a leg	Juda qimmatga tushmoq	Qo'l va oyoq narxiga tushmoq	Juda katta mablag' talab qiladigan narsa
Under the weather	sillasi qurigan	Ob-havo ostida	Kasal yoki charchagan bo'lish
Kill two birds with one stone	Bir o'q bilan ikki quyinni urmoq	Bir tosh bilan ikki qushni o'ldirmoq	Bir harakat bilan ikkita ishni bajarish
When pigs fly	Tuyani dumi yerga tekkanda	Cho'chqalar uchganda	Umuman amalga oshmaydigan yoki imkonsiz narsani ifodalaydi
Add fuel to the fire	Olovga yog'	Yong'inga benzin quyish	Allaqachon yomon bo'lgan holatni og'irlashtirish
Cry over spilled milk	O'tgan salovat	To'kilgan sut ustida yig'lamoq	Qaytarib bo'lmaydigan narsaga achinish befoйда ekanligini bildiradi

¹⁵⁵ Botir Elov va Ma'rufjon Amirqulov. Uzbeki-English Parallel Corpus Algorithm and Alignment Problem. – Central Asian Journal Of Literature, Philosophy And Culture. 2023. <https://www.researchgate.net/publication/372968220>

Frazeologik birliklarning struktural tahlili ularning sintaktik shakllarini o'rganadi. Ingliz va o'zbek tillarida frazeologik birliklar grammatik jihatdan turlicha bo'lishi mumkin.

– Inglizcha: “Make a decision” – bu frazeologik birikma oddiy so'zlashuvda ishlatiladi, lekin grammatik jihatdan biror “action” (harakat)ni ifodalovchi fe'lni talab qiladi.

– O'zbekcha: “Qaror qabul qilish” – bu ifoda ham “qaror” (nom) bilan ishlaydi, lekin o'zbek tilidagi frazeologik birikmalarni tuzishda aniq bir fe'l qo'llanadi.

Bu turdagi frazeologik birliklarning parallel korpusda aks etishi sinxron tilshunoslikda ko'riladi. Parallel korpus tarjima tilining struktur tuzilishidan tashqari so'z, ibora kollokatsiya, idiom va so'z birikmalarining muayyan kontekstdagi imkoniyatlarini kuzatishga ham xizmat qiladi. Ushbu korpus turlari zamonaviy tilshunoslikda, qiyosiy tilshunoslikda (ikki til strukturasini tahlil qilish maqsadida); tarjima sohasida (asliyat tildagi matnning kontekstual tarjimasini boshqa tillardan qidirish maqsadida); avtomatik tarjima (elektron lug'atda); leksikografiya sohalarida keng foydalanish mazkur soha tarmoqlarini kengayishiga imkon yaratadi. Bu esa parallel korpus yaratishning tizimlarining ilmiy va amaliy jihatlarini chuqurroq o'rganishni talab etadi. “Fraza birliklari **ierarxik sintaktik tuzilmalar** bo'lib, ular inson miyasida mavjud bo'lgan **tug'ma grammatik qoidalar** asosida shakllanadi. Fraza birliklarini segmentatsiya qilish **konstituentlik** (ya'ni tarkibiy qismlar) asosida amalga oshiriladi”¹⁵⁶. Ingliz va o'zbek tillaridagi frazeologik birliklar strukturaviy jihatdan ham farq qiladi. Ingliz tilida frazeologik birliklar ko'pincha fe'l+to'ldiruvchi yoki predikat+obyekt shaklida bo'ladi, o'zbek tilida esa bu birliklar ko'proq iboraviy yoki fe'l-bog'lama+ot ko'rinishida bo'ladi.

Inglizcha	Struktura turi	O'zbekch	Struktura turi
Spill the beans	Verb + Object	Sirni ochmoq	Qo'shma fel
A piece of cake	Noun + Predicative	Juda oson	Orttirma sifat darajasi
Hit the nail on the head	Verb + Adverb	Kalavani uchini topmoq	To'g'ri topmoq
Kick the bucket	Verb + Object	Vafot etmoq	Qo'shma fe'l
Bite the bullet	Verb + Object	Qiyinchilikka chidash	Qo'shma fe'l
Under the weather	Noun + Predicative	Sillasi qurigan	Qo'shma fe'l
Let the cat out of the bag	Verb + Object	Sirni oshkor qilmoq	Ot + qo'shma fe'l
Once in a blue moon	Noun + Predicative	Onda-sonda	Juft so'z
Cry over spilled milk	Verb + Object	O'tgan ishga salovat	Fe'l + ot + fe'l
Kill two birds with one stone	Verb + Object	Bir o'q bilan ikki quyonni o'ldirmoq	Ot + ko'makchi + fe'l
Cost an arm and a leg	Noun + Predicative	Juda qimmatga tushish	Sifat + fe'l
The ball is in your court	Verb + Adjective	Endi navbat sendan	Ravish + ot + olmosh
Add fuel to the fire	Verb + Object	Olovga yog' sepmoq	Ot + fe'l
When pigs fly	Verb + Adjective	Tuyani dumi yerga tekkanda	Ot + ot + fe'l

Frazeologik birliklarni segmentlash – bu matn ichida frazeologik iboralarni aniqlab, ularni oddiy so'zlardan farqli holda bir butun birlik sifatida ajratish jarayonidir. Segmentlash jarayoni bu iboralarning kontekstdagi semantik yaxlitligini saqlab qolish va ularni avtomatik tarzda tahlil

¹⁵⁶ **Noam Chomsky va Maurice Gross.** From The Concept Of Transformation In Harris And Chomsky To The Lexique-Grammaire Of Maurice Gross. – Sankt peterburg, 2011.

qilish imkonini beradi. Nilufar Abduraxmonova parallel korpuslarning leksik va semantik tadqiqotlarda ham foydaliligini ta'kidlaydi. "Ingliz tilidagi sodda gapshakllarini o'zbek tiliga avtomatik tarjima qilish namunasi orqali dasturiy ta'minotga kiritilayotgan bilimlar bazasiga lingvistik nuqtai nazardan yondashiladi. Chunki tarjima qilinayotgan tillar tabiiy tillardir. Ularning xususiyatlaridan kelib chiqib nazariy asoslarini berish orqali sifatli tarjimaga erishish mumkin bo'ladi"¹⁵⁷. Xulosa qilib aytganda, hozirda tarjimani avtomatlashtirishda parallel korpusdan foydalanish va uni amaliyotga tatbiq qilish natijasida bir qancha yutuqlarga erishilmoqda. Shunday ekan, qanday koprus yaratishdan qat'i nazar dastlab korpus yaratish uchun duch keladigan muammolarni chuqurroq o'rganish muhim va unga yechim topish uchun korpus turlarining tizimlarini o'zlashtirish lozim.

Foydalanilgan adabiyotlar:

1. Abdurahmonova N. Mashina tarjimasi: lingvistik asoslar. – T.: Akademnashr, 2012. – B. 21-25.
2. Gatiatullin, A., Suleymanov, D., Prokopyev, N., & Abdurakhmonova, N. (2022, November). "Turkic Morpheme": From the Portal to the Linguistic Platform. In World Conference Intelligent System for Industrial Automation (pp. 181-189)
3. Noam Chomsky, Maurice Gross. From The Concept Of Transformation In Harris And Chomsky To The Lexique-Grammaire Of Maurice Gross. – Sankt peterburg, 2011.
4. Karimov Rustam, Mengliyev Baxtiyor. The Role of the Parallel Corpus in Linguistics, the Importance and the Possibilities of Interpretation. – International Journal of Engineering and Advanced Technology (IJEAT). 2019.
5. https://tsuull.uz/sites/default/files/the_role_of_the_parallel_corpus_in_linguistics.pdf
6. Botir Elov va Ma'rufjon Amirqulov. Uzbeki-English Parallel Corpus Algorithm and Alignment Problem. – Central Asian Journal Of Literature, Philosophy And Culture. 2023. <https://www.researchgate.net/publication/372968220>

¹⁵⁷ Abdurahmonova N. Mashina tarjimasi: lingvistik asoslar. – T.: Akademnashr, 2012. 21-25-bet.

O'ZBEK VA RUS MILLIY KORPUSINING ARXITEKTURASI

Giyasova Dilrabo Turg'un qizi

O'ZMU Jurnalistika va o'zbek filologiyasi fakulteti
Filologiya va tillarni o'qitish: o'zbek tili yo'nalishi talabasi
E-mail: dilrabogiyasova01@gmail.com

Ilmiy rahbar: **A.R. Iskandarova**
f.f.n., dotsent

Annotatsiya: Maqolada o'zbek va rus milliy korpuslari tahlil qilingan. Ularning tuzilishi, rivojlanish bosqichlari haqida ma'lumot berilgan. So'nggi yillarda korpuslarni takomillashtirish borasida olib borilayotgan tadqiqotlarga to'xtalangan.

Kalit so'zlar: Parallel korpus, ta'limiy korpus, tezaurus, o'quv lug'atlari, token, lemma.

Аннотация: В статье анализируются узбекский и русский национальные корпуса. Приведена информация об их структуре и стадиях развития. Приведены примеры исследований, проведенных в последние годы по совершенствованию корпусов.

Ключевые слова: Параллельный корпус, образовательный корпус, тезаурус, образовательные словари, токен, лемма.

Abstract: The article analyzes the Uzbek and Russian national corpora. Information is provided about their structure and stages of development. Examples of research conducted in recent years on improving corpora are given.

Keywords: Parallel corpus, educational corpus, thesaurus, educational dictionaries, token, lemma.

XX asrning ikkinchi yarmidan boshlab amaliy tilshunoslikda korpus va korpus texnologiyalarini yaratishga katta e'tibor qaratildi. Dastlab 1963-yilda Braun universitetida ilk katta hajmli matnlarni o'z ichiga olgan *Brown korpusi* yaratildi. Keyingi yillarda *LOB (Lancaster Oslo/ Bergen)*, *Upsula rus tilining milliy korpusi*, *British National Corpus*, *Amerika milliy korpusi*, *Amerika ingliz tili korpusi*, *Rus tilining milliy korpuslari* yaratildi. Korpus milliy tilni saqlab qolish hamda uni raqamlashtirish imkoniyatini yaratadi. Dunyo tajribasida korpus yaratishning lingvistik, matematik va dasturiy tomonlari olimlarning tadqiqotlarida o'z ifodasini topgan. Korpus lingvistikasi bo'yicha jahonda V.Zaxarov, A.Sedov, A.Baranov, R.Potapova, V.Rikov, V.Martin, A.Laurens, A.Murakami, P.Tompson, Aksan Deniz, Yusup Aibaidulla va boshqa olimlar tadqiqotlar olib borgan. O'zbek korpus lingvistikasida B.Mengliyev, N.Abduraxmonova, Sh.Shahobiddinova, Z.Xolmonova, S.Karimov, L.Raupova, Sh.Hamroyeva, G.Toirova, J.Djumabayeva, G.Ergasheva, A.Eshmo'minov, Sh.Gulyamovalarning tadqiqotlari alohida ahamiyat kasb etadi.

Korpus lingvistikasi doirasida professor Nilufar Abduraxmonova rahbarligida olib borilayotgan tadqiqotlar natijasida O'zbek milliy korpusi yaratildi. Yozma va o'g'zaki matn namunalari bilan boyitilmoqda [<https://uzbekcorpus.uz/>]. Olimaning ta'kidlashicha, o'zbek tili elektron korpusining kompyuter modellari ikki xil yondashuvga asoslanadi:

- 1) korpus uchun yaratilgan tayyor lingvistik instrumentariylar va platformalar;
- 2) korpus taksonomiyasidan kelib chiqib, turli maqsadlarga mo'ljallangan kompyuter lingvistikasi metodlari orqali yaratilgan formal-funksional modellar [Abduraxmonova N. 2023].

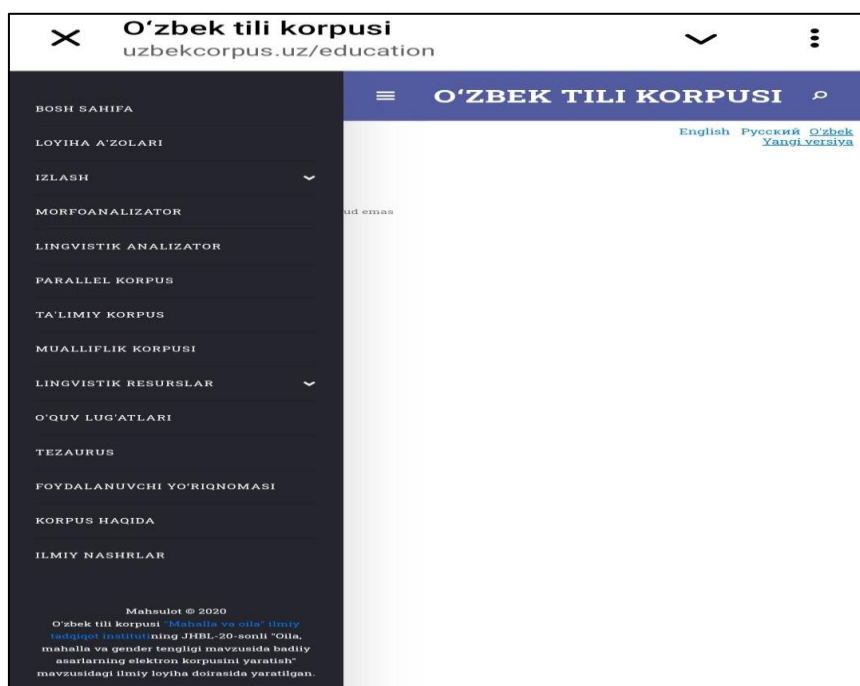
O'zbek korpusi o'zbek tilidagi 10 milliondan ortiq so'zlarni qamrab olgan. Joriy yilning fevral va mart oylari mobaynida N.Abduraxmonova va shogirdlari tomonidan korpusning yangi versiyasi yaratildi. Yangilik sifatida Parallel korpus tarkibiga turk hamda koreys tillari qo'shildi. O'zbek tilining ushbu elektron korpusi bir necha yillik olib borilgan tadqiqotlar, ilmiy hamkorlik

va loyihalardan olingan natijalar mahsuli hisoblanadi. O‘zbek tili korpusining joriy holatida o‘zbek tilida yaratilgan lingvistik lug‘atlar, web sahifalar, o‘zbek tilining morfologik ma’lumotlar bazasi, o‘quv adabiyotlar hamda turli janrdagi ilmiy, rasmiy va badiiy matnlar majmuasidan tashkil topgan.

Korpus – og‘zaki va yozma matnlar yig‘indisi sifatida, ma’lumotlarni boshqarish va foydalanish imkoniyatiga ega dasturiy ta’minotdir. Ushbu korpusning tarkibida ichki – subkorpuslar ham mavjud:

- Parallel korpus;
- Mualliflik korpusi;
- Ta’limiy korpus.

Shuningdek, turli lingvistik resurslar: o‘zbek tilining tezaurusi, ingizcha-o‘zbekcha fe’li frazemalar lug‘ati hamda boshqa lingvistik lug‘atlar o‘rin olgan. O‘zbek korpusidan ingliz, rus va o‘zbek tillarida foydalanish mumkin.



O‘zbek korpusida izlash bo‘limi orqali muayyan so‘zni token, lemma, birikma, konkordans bo‘yicha qidirish va tahlil qilish mumkin. Parallel korpus bo‘limidan o‘zbekcha-inglizcha, inglizcha-o‘zbekcha, o‘zbekcha-ruscha, ruscha-o‘zbekcha, o‘zbekcha-turkcha, turkcha-o‘zbekcha, o‘zbekcha-koreyscha, koreyscha-o‘zbekcha parallel matnlar kiritilgan.

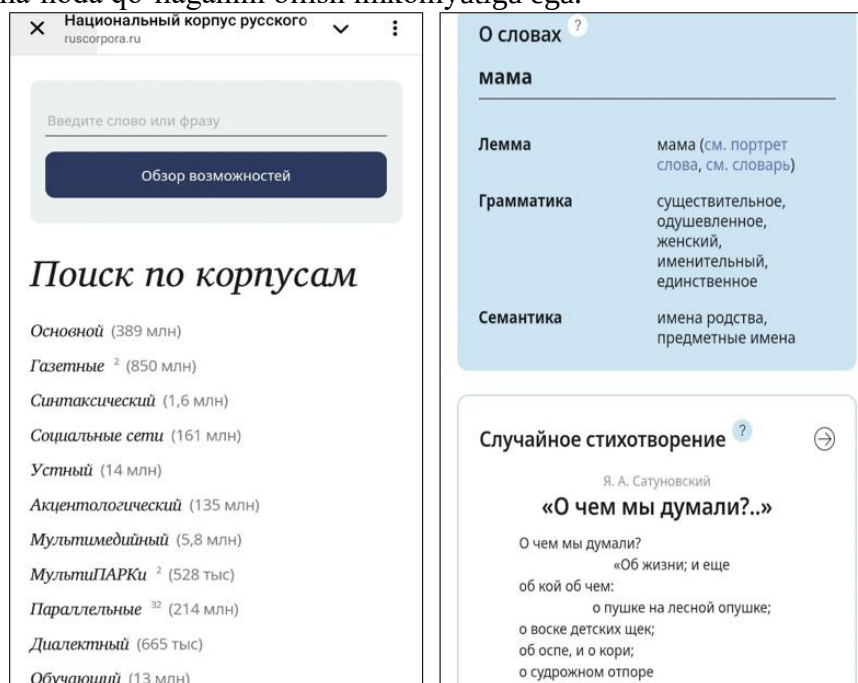
Parallel matnlar uchun segmentlash va lingvistik tahlil jarayonida Wordfast kabi instrumentlar samarali texnologik vosita sanaladi. Korpus tarkibdagi mualliflik korpusida so‘zni qo‘llangan matni, matn muallifi, uslubi, davri kabi ma’lumotlari bilan topish mumkin. Mualliflik korpusi bo‘yicha *Sh.Hamrayeva* ham izlanishlar olib borgan [Hamrayeva Sh. 2020].

Rus tili korpusining arxitekturasi ikki massivli yozma matnlardan iborat. Korpusning birinchi qismi XX asr o‘rtalari va XXI asr boshlarida yaratilgan matnlar, ikkinchi qismi XVIII asr oxiri - XX asr boshlariga tegishli yozma manbalarni o‘z ichiga olgan [<https://ruscorpora.ru>]. Uning kontentida rus tilining yozma adabiy tiliga doir zamonaviy badiiy proza janrlari va uning yo‘nalishlari, zamonaviy dramaturgiya, memuar va biografik adabiyotlar, jurnalistika va adabiy tanqid, gazeta jurnalistikasi va yangiliklar, ilmiy, ilmiy-ommabop va o‘quv materiallar, diniy va diniy-falsafiy matnlar, rasmiy ish va qonun hujjatlari, kundalik matnlar (shu jumladan nashr uchun mo‘ljallanmagan matnlar: shaxsiy yozishmalar, kundaliklar va boshqalar) o‘rin olgan [Сичи́навы Д.В. 2005].

Rus milliy korpusiga 2 milliard so‘z kiritilgan. Rus milliy korpusi XI asrning o‘ninchi yillaridan XXI asrgacha bo‘lgan davrdagi leksemalarni o‘z ichiga qamrab oladi. Bu korpusi

homiyisi Яндекс kompaniyasi hisoblanib, korpusni yaratishda V.V.Vinogradov nomidagi rus tili instituti, “Oliy ekonomika maktabi” milliy tadqiqot instituti, Sant-Peterburgdagi Lingvistik tadqiqot instituti, Axborot yetkazish muammolari instituti, Voronej davlat universitetining professor olimlari, talabalar, aspirant volontyorlari ishtirok etishgan. Rus milliy korpusi asosiy, sintaktik, gazeta, ijtimoiy tarmoq, og‘zaki, aksent, multimedia, multipark, parallel, dialektal, ta’limiy, 2dan 15gacha, epik, rus klassikasi, tarixiy, panxronik subkorpuslarni o‘z ichiga qamrab oladi. Parallel korpus 32ta tilda ishlay oladi. Foydalanuvchi o‘ziga kerakli subkorpusni tanlash orqali ko‘pgina ma’lumotlarga ega bo‘ladi. Multimedia joylash bo‘yicha rus korpusi boshqa korpuslarga qaraganda yaxshi rivojlanib kelmoqda. Multimedia subkorpusiga turli xil video qo‘llanmalar joylangan. Bu video qo‘llanmalar razmetkalangan bo‘lib, foydalanuvchiga turli qulayliklarni yaratadi. Masalan, “**Yangiliklar**” degan joyida har oyda yaratilgan yangiliklar kiritib boriladi.

Joriy yilning mart oyida rus klassikasi subkorpusiga 7,5 million so‘z joylandi. Bunda Dostoyevskiy, Nekrasov, Turgenov va boshqa yozuvchilarning asarlari kiritilgan. Endi korpus orqali 9ta muallif asarlaridagi sinonim so‘zlarni aniqlash mumkin. Korpusdan foydalanuvchilar bu yozuvchilar o‘z asarlarida sinonim so‘zlarni qay darajada bir-birlariga yaqin va bir-birlaridan uzoq bo‘lgan ma’noda qo‘llaganini bilish imkoniyatiga ega.



Национальный корпус русского
ruscorpora.ru

Введите слово или фразу

Обзор возможностей

Поиск по корпусам

- Основной (389 млн)
- Газетные ² (850 млн)
- Синтаксический (1,6 млн)
- Социальные сети (161 млн)
- Устный (14 млн)
- Акцентологический (135 млн)
- Мультимедийный (5,8 млн)
- МультиПАРКи ² (528 тыс)
- Параллельные ³² (214 млн)
- Диалектный (665 тыс)
- Обучающий (13 млн)

О словах ?

мама

Лемма	мама (см. портрет слова, см. словарь)
Грамматика	существительное, одушевленное, женский, именительный, единственное
Семантика	имена родства, предметные имена

Случайное стихотворение ?

Я. А. Сатуновский

«О чем мы думали?..»

О чем мы думали?
 «Об жизни; и еще
 об кой об чем:
 о пушке на лесной опушке;
 о воске детских щек;
 об оспе, и о кори;
 о судрожном отпоре

Rus milliy korpusining “**Imkoniyatlar sharhi**” qidiruv tizimiga biror bir so‘z yozilsa, qanday qulayliklar mavjud ekanligi haqida ma’lumot beradi. Masalan, “**Ona**” leksemasi kiritilsa, shu leksemaning lemmasi, grammatikasi va semantikasi haqida ma’lumotlar beriladi. Bundan tashqari, ushbu leksema qaysi she’rlarda mavjudligi ham ko‘rsatiladi. Panxronik subkorpus orqali shu so‘z qaysi yillarda qo‘llanilganligi grafik holda kollokatsiyasi bilan birga chiqarib beriladi. Kollokatsiyadan “**Ona**” leksemasi bilan birga qanday so‘zlar kelishi mumkinligini bilib olish mumkin. Aksent subkorpusidan rus tilini o‘rganuvchilar foydalanishlari mumkin. Bunda foydalanuvchi ma’lum bir leksemani yozadi va matnlar chiqarib beriladi. Matnlarda keltirilgan va qidiruvga berilgan so‘z ham urg‘u belgisi bilan yozilgan bo‘ladi. Ushbu urg‘u orqali foydalanuvchi so‘zni to‘g‘ri o‘qishni va talaffuz qilishni o‘rganib oladi.

Yildan yilga korpus lingvistikasi rivojlanib bormoqda. Tabiiy tilni qayta ishlashga e’tibor qaratilmoqda. Korpuslarga milliy tilning grammatikasida mavjud qonun-qoidalar kiritilmoqda. Bu orqali tillar yanada rivojlanib, avloddan avlodga to‘liq va oson yetib borishiga erishilmoqda. Korpuslar orqali tilning leksik, semantik, grammatik jihatdan o‘rganishimiz mumkin va chet elliklarga o‘rgatishga muaffaq bo‘lamiz. O‘zbek va rus milliy korpuslari ham so‘nggi yillarda juda

katta yutuqlarga erishmoqda. O‘zbek korpusi yaratilganiga ko‘p vaqt bo‘lmagan, shunga qaramasdan, turli uslublarga doir matnlar joylanib, uni mukammallashtirish borasida ilmiy izlanishlar olib borilmoqda.

Adabiyotlar ro‘yxati:

1. Abduraxmonova N. Korpus lingvistikasi. T.:GlobeEdit, 2023.
2. <https://uzbekcorpus.uz/education>
3. <https://ruscorpora.ru>
4. Д.В.Сичинавы. Национальный корпус русского языка: очерк предыстории// Национальный корпус русского языка: 2003-2005. Результаты и перспективы. Москва, 2005.
5. Хамраева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари (монография). Глобэдит, 2020.

ULUGBEK HAMDAM IJODIDA LINGVOSTATISTIK TAHLIL ("TO'LIN OY" QISSASI ASOSIDA)

Xolbekova Dilrabo Xusniddin qizi

O'zbekiston Milliy universiteti
Kompyuter lingvistikasi va amaliy
tilshunoslik kafedrası o'qituvchisi

Annotatsiya: Ushbu maqolada Ulug'bek Hamdamning "To'lin oy" qissasi lingvostatistik metodlar asosida tahlil qilindi. AntConc dasturi yordamida asar matnining so'z chastotasi, leksik boyligi va qahramonlar ismlarining uchrash chastotasi tahlil etildi. Tahlil natijalari qissaning badiiy, poetik va semantik jihatdan izchil va mazmunli tuzilganini ko'rsatdi.

Kalit so'zlar: lingvostatistika, To'lin oy, Ulug'bek Hamdam, leksik boylik, so'z chastotasi, poetik matn, adabiy tahlil.

Аннотация: В данной статье произведён лингвостатистический анализ повести Улугбека Хамдама «Полная луна». С помощью программы AntConc были исследованы частотность слов, лексическое богатство и частота упоминания имён персонажей. Результаты анализа показали, что повесть имеет последовательную и содержательную художественную, поэтическую и семантическую структуру.

Ключевые слова: лингвостатистика, Полная луна, Улугбек Хамдам, лексическое богатство, частотность слов, поэтический текст, литературный анализ.

Abstract: This article presents a linguostatistical analysis of Ulug'bek Hamdam's novella To'lin oy (Full Moon). Using AntConc software, the frequency of words, lexical richness, and the occurrence frequency of character names were analyzed. The results show that the novella is structured in a coherent and meaningful way in terms of artistic, poetic, and semantic aspects.

Keywords: linguostatistics, Full Moon, Ulug'bek Hamdam, lexical richness, word frequency, poetic text, literary analysis.

Bugungi kunda adabiy matnlarni tahlil qilish jarayonida zamonaviy lingvistik metodlardan, xususan, lingvostatistik yondashuvdan foydalanish samarali natijalar bermoqda. Lingvostatistik tahlil matn tarkibidagi so'zlar, ularning shakllari, takrorlanish chastotasi, leksik va grammatik birliklarning taqsimlanish qonuniyatlarini aniqlash, shuningdek, badiiy uslub xususiyatlarini matematik ko'rsatkichlar asosida tahlil etish imkonini beradi.

Lingvostatistik tadqiqotlar tilshunoslikda 20-asrning o'rtalaridan boshlab faol rivojlana boshladi. Xususan, rus va o'zbek tillarining qiyosiy fonetikasi sohasida dastlabki statistik izlanishlar V.M.Popov tomonidan amalga oshirilgan bo'lib, uning "O'zbek va rus tillarini qiyosiy o'rganishda statistik metodni qo'llash" nomli maqolasi (1949) bu boradagi ilk tadqiqotlar sirasiga kiradi. Ushbu maqolada har ikki tildagi tovushlarning nutqdagi o'rni, ishlatilish chastotasi, yumshoqlik va qattqlik xususiyatlari statistik tahlil asosida aniqlangan. [Yo'ldoshev M., Muhamedova S., Saparniyazova M., 2021: 160]

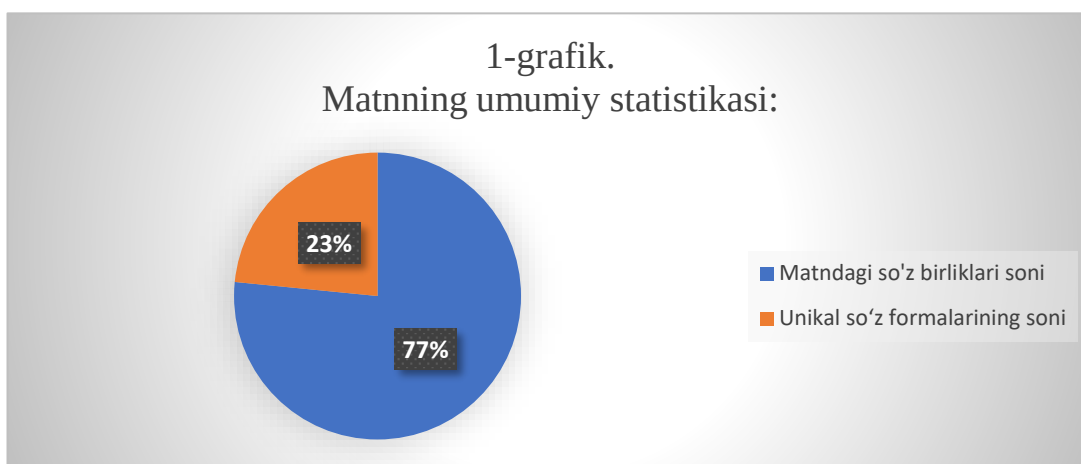
Shuningdek, o'zbek adabiy tilining leksik tizimini statistik jihatdan o'rganishga qaratilgan tadqiqotlar ham ahamiyatli natijalar bermoqda. Tilshunos olim B.Bafoyev o'zining Alisher Navoiy asarlari leksikasiga bag'ishlangan monografiyasida ushbu adib asarlari tilining statistik tahlil natijalarini alohida bob sifatida ko'rib chiqqan. Monografiyaning "So'z boshi" qismida esa Alisher Navoiy asarlari tilini statistik jihatdan o'rganish yuzasidan olib borilgan tadqiqotlar, e'lon qilingan lug'atlar hamda qo'lyozma lug'atlardagi izohlangan so'zlar miqdori haqida ma'lumotlar berilgan. [Yo'ldoshev M., Muhamedova S., Saparniyazova M., 2021: 161]

Adabiy matnlarni lingvostatistik jihatdan o'rganish nafaqat yozma adabiyot, balki xalq og'zaki ijodi namunalari uchun ham muhim ahamiyat kasb etadi. Bu borada professor Dilbar

Bazarovna Urinbayeva tomonidan tayyorlangan “Xalq og‘zaki ijodi asarlari matnining janriylisoniy va lingvostatistik tadqiqi” nomli monografiya e‘tiborga molikdir. Tadqiqot obyekti sifatida o‘zbek xalq og‘zaki ijodining doston, maqol, topishmoq, ertak va qo‘shiq matnlari tanlab olingan bo‘lib, ayniqsa “Alpomish”, “Ravshan”, “Topishmoqlar”, “O‘zbek xalq maqollari”, “O‘zbek xalq ertaklari” va “Boychechak” kabi asarlar tahlil qilingan. Ushbu matnlar asosida tuzilgan “O‘zbek xalq dostonlari tilining alfavitli-chastotali lug‘ati” (2006), “O‘zbek xalq dostonlari tilining chastotali lug‘ati” (2006) hamda “O‘zbek xalq dostonlarining chastotali izohli lug‘ati” (2008) singari leksikografik tadqiqotlar monografiya yozilishida asosiy manba sifatida xizmat qilgan. [Urinbayeva D., 2022: 8]

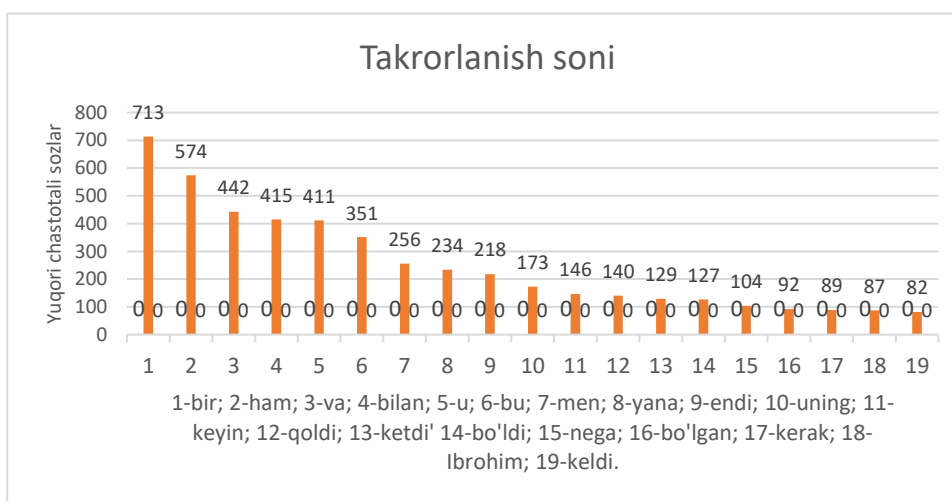
Zamonaviy lingvistik tadqiqotlarda lingvostatistik tahlil, ayniqsa, kompyuter lingvistikasining rivojlanishi bilan yanada keng qo‘llanilmoqda. N.Abdurahmonovanning ilmiy ishida qayd etilishicha, o‘zbek tilida mashina tarjima sohasida tadqiqotlarning yetarli darajada olib borilmayotgani sababli, tarjima jarayonini takomillashtirish maqsadida o‘zbek tilining lingvistik ma‘lumotlar bazasini yaratish zarurligi ta‘kidlangan. [Абдурахмонова Н., 2018: 4]

Ushbu maqolada esa zamonaviy lingvostatistik tahlil imkoniyatlaridan foydalanilgan holda Ulug‘bek Hamdam qalamiga mansub “To‘lin oy” qissasi AntConc dasturi yordamida lingvostatistik jihatdan tahlil qilindi.



Leksik boylik koeffitsiyenti: 30.6%. Bu ko‘rsatkich matnning uslubiy va semantik jihatdan boy ekanini tasdiqlaydi.

Matnda quyidagi so‘zlar eng ko‘p takrorlangan:



2-grafik. Yuqori chastotali so‘zlar statistikasi:

Fe'llar: qoldi, ketdi, keldi, chiqib, bo'ldi, so'radi, ko'rdi.

Ravish va bog'lovchilar: yana, endi, keyin, nega, deya.

Ot va sifatlar: oy, yo'l, hayot, quvonch, xayol, qorong'ulik.

Bu ko'rsatkichlar matnda harakatning dinamikasi va qahramonlarning hissiy holatlari til vositalari orqali qanday ifodalanganini ko'rsatadi.

Qahramonlar markazi: So'z chastotasi asosida Ibrohim, Kozimjon, Zulayho va Hasan ismlarining matnda ko'p uchrashi ushbu obrazlarning qissa syujeti rivojidadagi markaziy rolini ko'rsatadi.

Matn poetikasi: "To'lin oy" qissasida ramz va poetik obrazlar tili boy. "Oy", "yolg'izlik", "ko'z", "umid", "tun", "quyosh" so'zlari matnda hissiy-falsafiy qatlam hosil qilib, muallif tilining poetik jihatlarini ochib beradi.

Ulug'bek Hamdamning "To'lin oy" qissasini lingvostatistik tahlil qilish natijasida aniqlanishicha, asar matni murakkab va ma'noli struktura asosida qurilgan bo'lib, muallif so'z tanlovi orqali qahramonlarning ichki kechinmalarini, tasvirlangan voqealar manzaralarini va ularning psixologik holatini yuksak mahorat bilan ifodalagan.

Qissadagi so'zlar chastotasi orqali aniqlangan birliklar, ayniqsa, qahramonlar ismlari, ramziy va mavzuli so'zlar muallif tilining badiiy imkoniyatlarini to'liq namoyon qiladi. Ayniqsa, "oy", "ko'z", "yolg'izlik" kabi ramziy so'zlar asar semantik markazini shakllantirib, muallifning falsafiy fikrlashini aks ettiradi.

Shuningdek, yuqori chastotali fe'llar voqealar rivoji va harakat dinamikasini aks ettirishda xizmat qilgan, olmoshlar esa hikoya subyektivligini ta'minlagan. Bu kabi tahlillar, matn mazmunini statistik faktlar orqali yanada chuqur anglash imkonini beradi.

Qissadagi yuqori leksik boylik ko'rsatkichi, mavzu va personajlarning semantik yaqinligi, uslubiy yondashuvlarning xilma-xilligi Ulug'bek Hamdam ijodining o'ziga xosligini va adabiy mahoratini tasdiqlaydi.

Lingvostatistik tahlil natijalari shuni ko'rsatadiki, "To'lin oy" qissasi nafaqat adabiy-badiiy, balki ilmiy-lingvistik nuqtayi nazardan ham mukammal asar sifatida baholanishi mumkin. Ushbu tahlil matnning poetik va semantik qatlamlarini ochib, uning asosiy g'oyaviy yondashuvlarini, muallif falsafasini yanada chuqur tushunishga imkon beradi.

Foydalanilgan adabiyotlar:

1. Urinbayeva D.B "Xalq og'zaki ijodi asarlari matnining janriy-lisoniy va lingvostatistik tadqiqi" "Qamar media" nashriyoti Toshkent – 2022
2. M.Yo'ldoshev, S.Muhammedova, M.Saparniyozova "Matn lingvistikasi" Toshkent «Ishonchli hamkor» 2021
3. Абдурахмонова Н. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти: Филол. фан. фалсафа фан. д-ри (phd). дисс. – Тошкент, 2018. –Б. 4.
4. Мухамедов С. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дис.... канд. филол. наук. -Ташкент, 1980. – С.25.20
5. Ulug'bek Hamdam "To'lin oy" qissasi

ELEKTRON MATN KORPUSLARI VA ULARNING DIALEKTOLOGIK TADQIQOTLARDAGI AHAMIYATI (Fargʻona viloyati shevalari misolida)

Vosiljonov Azizbek

Fargʻona davlat universiteti oʻqituvchisi

Annotatsiya: Ushbu maqolada hozirgi zamon lingvistik tadqiqotlarida elektron matn korpuslarining tutgan oʻrni, xususan, dialektologiya sohasida ularning qoʻllanilish imkoniyatlari tahlil qilinadi. Fargʻona viloyati shevalari misolida dialektal korpus yaratish jarayoni, uning tarkibiy qismlari, metodik yondashuvlari va ilmiy ahamiyati keng yoritiladi. Shuningdek, zamonaviy korpus lingvistikasi yutuqlari asosida hududiy shevalarni tizimli oʻrganish, ularning fonetik, leksik va grammatik xususiyatlarini aniqlash orqali oʻzbek tilining dialektal boyligini chuqur tahlil qilish imkoniyatlari asoslab beriladi.

Kalit soʻzlar: korpus tilshunosligi, dialektal korpus, Fargʻona shevalari, lingvistik tadqiqot, transkripsiya, metarazmetka.

Kirish

Soʻnggi yillarda axborot texnologiyalari taraqqiyoti tilshunoslik fanining barcha tarmoqlarida, xususan, korpus tilshunosligi va dialektologiyada yangi imkoniyatlar yaratmoqda. Til hodisalarini zamonaviy usullar bilan oʻrganish, ularni statistik va texnologik jihatdan tahlil qilish bugungi lingvistik tadqiqotlarning ajralmas qismiga aylangan. Ana shunday metodik yondashuvlardan biri bu – elektron matn korpuslari asosida til materiallarini tizimli tahlil qilishdir. Til korpuslari – bu maʼlum qoidalar asosida toʻplangan, tartiblangan, tuzilgan va maxsus razmetka (belgilash) berilgan matnlar bazasi boʻlib, u kompyuter orqali tahlil qilish uchun moʻljallangan. Ular zamonaviy lingvistika, ayniqsa, amaliy va hisobiy lingvistika uchun asosiy manbalardan biri hisoblanadi.

Asosiy qism

Korpus tilshunosligi, eng avvalo, katta hajmdagi real til materiallarini kompyuter orqali tahlil qilishga asoslanadi. Korpuslar oʻz mazmuniga koʻra turli turlarga ajraladi: umumiy til korpuslari, ixtisoslashgan (sohaviy) korpuslar, tarixiy korpuslar, bolalar nutqi korpuslari, dialektal korpuslar va boshqalar. Aynan dialektal korpuslar esa maʼlum hududda yashovchi aholining ogʻzaki va yozma nutq namunalarini oʻz ichiga olgan, strukturaviy va semantik jihatdan boyitilgan materiallar toʻplamidir. Bunday korpuslar orqali dialektlarning fonetik, morfologik, leksik, sintaktik xususiyatlarini keng qamrovda tahlil qilish, ularni adabiy til bilan solishtirish, oʻzaro qiyoslash mumkin.

Fargʻona viloyati shevalari rang-barangligi bilan ajralib turadi. Ushbu hududda yashovchi aholining ogʻzaki nutqida uchraydigan fonetik almashinuvlar, urgʻu tizimi, morfologik shakllar va leksik birliklar boshqa hududlardan farq qiladi. Bu esa ularni alohida oʻrganishni, lingvistik jihatdan tahlil qilishni talab etadi. Dialektal korpuslar yaratish orqali bu xususiyatlar tizimli tarzda hujjatlashtiriladi va kelgusida tahlil qilish uchun elektron manba sifatida saqlanadi.

Dialektal korpus yaratish jarayonining birinchi bosqichi – til materiallarini yigʻishdan boshlanadi. Bu jarayonda suhbatlar yozib olinadi, mavjud yozma manbalar (xatlar, esdaliklar, hikoyalar) toʻplanadi va ular transkripsiya qilinadi. Transkripsiya – bu matnni fonetik yoki fonematik shaklda yozib chiqish boʻlib, dialektal oʻziga xosliklarni aniq ifodalashda muhim ahamiyatga ega. Keyingi bosqichda matnlarga morfologik, sintaktik, semantik va boshqa razmetkalar qoʻyiladi. Ushbu razmetkalar korpusni kompyuter dasturlari orqali avtomatik tahlil qilish imkonini beradi.

Fargʻona viloyati shevalari asosida tuziladigan dialektal korpuslar bir necha asosiy funksiyalarni bajaradi. Birinchidan, u hududiy shevalarni hujjatlashtirish va saqlashda muhim vositadir. Ikkinchidan, ularni zamonaviy lingvistik tahlil usullari orqali oʻrganish imkonini beradi.

Uchinchi jihat esa amaliy foydalanish – ya’ni til o’rgatish, tarjima tizimlari, ovoz tanish dasturlari, sun’iy intellektga asoslangan til texnologiyalarini yaratishda foydalanish mumkin.

Shuningdek, dialektal korpuslar lingvistik tadqiqotlar doirasidan tashqarida ham muhim rol o’ynaydi. Masalan, ular madaniy merosni asrash, tilshunoslikka oid ta’lim dasturlarini boyitish, regional til siyosatini shakllantirish, toponimika va etnolingvistika kabi sohalarda ham qo’llanilishi mumkin. Ayniqsa, yosh avlodni o’z hududiy shevasi bilan tanishtirish, ularning til boyligini oshirishda dialektal korpuslar muhim vosita bo’lib xizmat qiladi.

Xulosa

Xulosa qilib aytganda, elektron matn korpuslari, ayniqsa, dialektal korpuslar bugungi kunda lingvistika fanining eng muhim yo’nalishlaridan biriga aylangan. Ular yordamida til hodisalari keng ko’lamda, chuqur tahlil qilinadi va tilning tabiiy holatini ilmiy asosda o’rganish imkoniyati yaratiladi. Farg’ona viloyati shevalari asosida yaratiladigan dialektal korpuslar o’zbek tilining dialektal boyligini tizimli o’rganish, uning rivojlanish tendensiyalarini aniqlash hamda hududiy shevalarning zamonaviy texnologiyalar orqali hujjatlashtirilishini ta’minlaydi. Bu esa tilni nafaqat o’rganish, balki uni saqlab qolish va rivojlantirishda ham muhim qadamdir.

Foydalanilgan adabiyotlar ro’yxati

1. Abdurahmonova N. Kompyuter lingvistikasi. – Toshkent: Nodirabegim, 2021. –394 b.
2. Turdialiyev A. Denov shevasini areal o’rganish va til korpusiga joylashtirish: Filol. fan. bo’yicha falsafa doktori (PhD) avtoref. – Toshkent, 2024.
3. Abdurakhmonova, N., & Urdishev, K. (2019). Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL), 6(1-2019), 131-7.
4. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All) (pp. 1-4).
5. Azizbek, V. (2024). JAHON KORPUS LINGVISTIKASIDA DIALEKTAL KORPUSLAR. «CONTEMPORARY TECHNOLOGIES OF COMPUTATIONAL LINGUISTICS», 2(22.04), 71-74.
6. <https://uzbekcorpus.uz/>

NORMATIV-HUQUQIY HUJJATLARNING HUQUQIY TIZIMDAGI O'RNI

Topildiyeva Feruza

O'zMU Kompyuter lingvistikasi
va amaliy tilshunoslik kafedrası
katta o'qituvchisi

Erkinova Sevinch Arslonovna

O'zMU, Jurnalistika va o'zbek
filologiyasi fakulteti 4-kurs talabasi

Annotatsiya: Mazkur maqolada normativ-huquqiy hujjatlarning huquqiy tizimdagi tutgan o'rni, ularning davlat boshqaruvidagi ahamiyati hamda jamiyat hayotini tartibga solishdagi roli yoritilgan. Shuningdek, normativ-huquqiy hujjatlarning turlari, ularning ierarxiyasi va huquqiy kuchi tahlil qilingan. Maqolada normativ-huquqiy hujjatlarning amaliyotdagi ahamiyati va ularni takomillashtirish zaruriyati haqida ham fikr yuritilgan.

Kalit so'zlar: Normativ-huquqiy hujjat, huquqiy tizim, huquq manbalari, qonun, qaror, ierarxiya, huquqiy tartibot, normativ baza.

Abstract: This article discusses the role of regulatory legal acts in the legal system, their importance in public administration and their role in regulating the life of society. It also analyzes the types of regulatory legal acts, their hierarchy and legal force. The article also discusses the importance of regulatory legal acts in practice and the need to improve them.

Keywords: Regulatory legal act, legal system, sources of law, law, decision, hierarchy, legal order, regulatory framework.

Аннотация: В статье рассматривается роль нормативных правовых актов в системе права, их значение в государственном управлении, а также их роль в регулировании общественной жизни. Также были проанализированы виды нормативных правовых актов, их иерархия и юридическая сила. В статье также рассматривается значение нормативных правовых документов в практике и необходимость их совершенствования.

Ключевые слова: Нормативно-правовой документ, правовая система, источники права, закон, решение, иерархия, правовой порядок, нормативная база.

Lingvistika, ya'ni tilshunoslik¹⁵⁸, insoniyatning eng muhim ijtimoiy vositalaridan biri bo'lib, har bir jamiyatda tilning roli va ahamiyati katta. Til nafaqat kundalik muloqotda, balki huquqiy va ijtimoiy sohalarida ham muhim ahamiyat kasb etadi. Shu boisdan, normativ-huquqiy hujjatlar tilni tartibga solish va amaldagi huquqiy tizimning samarali ishlashiga xizmat qiluvchi asosiy vosita bo'lib xizmat qiladi. Ushbu maqolada normativ-huquqiy hujjatlar va ularning lingvistik jihatlari, shuningdek, huquqiy tilning jamiyatdagi o'rni haqida so'z yuritimiz.

Normativ-huquqiy hujjatlar huquqiy tizimda muhim rol o'ynaydi, chunki ular davlat tomonidan belgilangan me'yor va qoidalarining amalda qanday qo'llanilishini aniqlaydi. Har bir normativ-huquqiy hujjat o'zining tili va uslubiga ega bo'lib, bu til va uslub huquqiy mazmunning aniq va tushunarli bo'lishini ta'minlashga xizmat qiladi. Misol uchun¹⁵⁹ konstitutsiyava huquqiy til haqida so'z yuritadigan bo'lsak, konstitutsiya, davlatning asosiy huquqiy hujjati bo'lib uning tili shakl-shubhasiz tilshunoslar uchun alohida qiziqish uyg'otadi. Konstitutsiya tili aniq va ishonchli bo'lishi kerak, chunki u davlatning bosh huquqiy asoslarini belgilaydi.

¹⁵⁸ Abdullayev A. Huquq nazariyasi va amaliyoti. "Toshkent; "Adolat" nashriyoti, 2020

¹⁵⁹ O'zbekiston Respublikasi "Normativ -huquqiy hujjatlar to'g'risida"gi qonuni - (Qabul qilingan sana; 2010-yil 14-dekabr/ oxirgi o'zgartirish bilan)

Qonunlarning tilini misol sifatida oladigan bo'lsak, ular rasmiy va juda aniq bo'lishi kerak. Qonunlarni tayyorlashda tilshunoslar, ayniqsa, leksik sintaktik va semantik jihatlarni hisobga olishadi, chunki bu hujjatlar fuqarolarning huquqiy holatini aniqlab beradi.

Prezident farmonlari va qarorlari¹⁶⁰ ham aniq va ravshan huquqiy tilga ega bo'lishi kerak. Ularning lingvistik xususiyatlari, odatda, ma'lum bir siyosiy yoki ijtimoiy masalalarni tartibga solish maqsadida ishlab chiqariladi.

Bunga yana misol bo'lib normativ-huquqiy aktlarni olishimiz mumkin. Ular maxsus idoralar va davlat organlari tomonidan ishlab chiqariladi. Bu turdagi hujjatlarning tili leksik jihatdan boy va ba'zida murakkab bo'lishi mumkin, chunki ular juda aniq me'yorlarni va normalarni belgilaydi.

Undan tashqari, huquqiy tilshunoslik degan termin ham mavjudki, u huquq va til me'yorlari va nutq uslublarini bir tizimga solgan holda o'rganadi. Qolaversa, huquqiy hujjatlarning aniq va tushunarli bo'lishi uchun o'ziga xos huquqiy terminologiya talab qilinadi. Har bir sohaning, jumladan, huquqiy sohaning o'ziga xos terminlari mavjud bo'lib ular jamiyatda muhim rol o'ynaydi.

Lingvistika va huquq o'rtasidagi bog'lanish jamiyatda huquqiy muvozanatni ta'minlashda katta ahamiyatga egadir.

Normativ-huquqiy hujjatlar¹⁶¹ va huquqiy hujjatlarning o'rganilishi, tilshunoslarga huquqiy tizimni tushunish va huquqiy matnlarning aniq ifodalashini ta'minlash imkonini beradi. Huquqiy tilning murakkabligi va aniq talablari, lingvistik yondashuvni va terminologiyani talab qiladi. Mavjud adabiy tilning yuridik asoslarni o'rganuvchi yuridik til sifatidagi o'rni, normativ hujjatlarning huquqiy tizimdagi shakllantirish jarayonini belgilab beradi.

Foydalanilgan adabiyotlar:

1. Mirziyoyev Sh.M. "Oliy Majlisga murojaatnoma", -Toshkent; O'zbekiston Respublikasi Prezident matbuot xizmati, 2023
2. O'zbekiston Respublikasi "Normativ -huquqiy hujjatlar to'g'risida"gi qonuni - (Qabul qilingan sana; 2010-yil 14-dekabr/ oxirgi o'zgartirish bilan)
3. Abdullayev A. Huquq nazariyasi va amaliyoti. "-Toshkent; "Adolat" nashriyoti, 2020
4. Rahmonqulov D. "Normativ hujjatlarning turlari va ularning ijtimoiy hayotdagi roli". - "Yuridik fanlar jurnali", 2022-yil №3

¹⁶⁰ Mirziyoyev Sh.M. "Oliy Majlisga murojaatnoma", -Toshkent; O'zbekiston Respublikasi Prezident matbuot xizmati, 2023

¹⁶¹ Rahmonqulov D. "Normativ hujjatlarning turlari va ularning ijtimoiy hayotdagi roli". - "Yuridik fanlar jurnali", 2022-yil №3

FHDYO HUJJATLARI LOYIHALARINO LINGVISTIK EKSPERTIZADAN O'TKAZISH

Topildiyeva Feruza

O'zMU Kompyuter lingvistikasi
va amaliy tilshunoslik kafedrası
katta o'qituvchisi

Erkinova Sevinch Arslon qizi

O'zMU, Jurnalistika va o'zbek
filologiyasi fakulteti 4-kurs talabasi

Annotatsiya: Mazkur maqolada FHDYO (Fuqarolik holati dalolatnomalarini yozish) hujjatlari loyihalarini lingvistik ekspertizadan o'tkazishning nazariy va amaliy jihatlari yoritilgan. Lingvistik ekspertizaning maqsadi – hujjat matnining huquqiy aniqligi, stilistik aniqligi va umumxalq til me'yorlariga mosligini tekshirishdir. Maqolada ekspertiza jarayonida e'tibor qaratilishi lozim bo'lgan lingvistik mezonlar, xatoliklar va ularni tuzatish yo'llari misollar bilan ochib beriladi. Shuningdek, FHDYO hujjatlarining o'ziga xos til xususiyatlari ham tahlil qilinadi.

Kalit so'zlar: yuridik hujjat, zaks, ariza, yo'llanma, yo'riqnoma, kishi faktori, ma'lumotnoma, qaror, ilova, ma'lumotnoma matni, huquqiy-normativ hujjat, lingvistik ekspertiza

Abstract: This article explores the theoretical and practical aspects of conducting linguistic expertise of Civil Registry Office (FHDYO) document drafts. The main goal of the expertise is to ensure the legal and stylistic accuracy of the text, as well as its compliance with standard language norms. The article outlines key linguistic criteria to be considered during the review process, discusses common errors, and provides examples of how to correct them. It also highlights specific language features typical for FHDYO documents.

Keywords: Legal document, Civil Registry Office (ZAGS), Application, Referral, Guideline / Instruction, Human factor, Certificate / Information note, Decision / Resolution, Annex / Attachment / Appendix, Text of the certificate, Legal and regulatory document, Linguistic expertise

Аннотация: В данной статье рассматриваются теоретические и практические аспекты проведения лингвистической экспертизы проектов документов ЗАГС. Цель экспертизы — проверка юридической и стилистической точности текста, а также его соответствия нормам литературного языка. В статье подробно анализируются лингвистические критерии, на которые следует обращать внимание при экспертизе, типичные ошибки и пути их устранения. Также рассматриваются языковые особенности документов ЗАГС.

Ключевые слова: Юридический документ, ЗАГС (Запись актов гражданского состояния), Заявление, Направление, Руководство, Человеческий фактор, Справка / Информационная записка, Решение / Постановление, Приложение, Текст справки, Правовой нормативный документ, Лингвистическая экспертиза

Oila qurish hayotdagi muhim bosqichlardan biridir va bu jarayonni osonlashtirish uchun turli yuridik hujjatlar talab etiladi. Zaks arizalariga oid yo'llanma-bu talaba yoki yoshlar uchun oila qurish va ro'yxatdan o'tish jarayonida zarur hujjatlarni to'plash bo'yicha yo'riqnoma hisoblanadi. Ammo, ba'zida bu yo'llanmalarda ma'lumotlar yangilanmaydi. Buning natijasida xatoliklar yuzaga keladi va yoshlar uchun noqulayliklarni keltirib chiqaradi. Ayniqsa, bu turdagi hujjatlarda imloviy xatoliklar ko'p kuzatilmoqda. Zaks hujjatlaridagi imloviy xatoliklar katta

qo'shimcha muommoni keltirib chiqarmasa ham, kichik bir e'tiborsizlik xalqning savodsizlik darajasini oshishiga olib kelishi tabiiy.

Zaks arizalari va yo'llanmalaridagi imloviy xatoliklar turli sabablarga ko'ra yuzaga keladi. Masalan, kompyuter tizimlari yoki avtomatik ariza to'ldirishda, xodimlarning e'tiborsizligi (arizalar tayyorlash jarayonida kishi faktori ham katta rol o'ynaydi), hujjatlarning ba'zi versiyalarida imloviy qoidalarga rioya qilmaslik yoki yangilanishlar vaqtida e'tibordan chetda qoldirish natijasida eski shakllar va imlo xatoliklari saqlanib qolishi mumkin.¹⁶²

Yaqinda oila qurishni niyat qilgan ikki yosh zaksga ariza topshirganida, ularga taqdim etilgan bir nechta yo'llanmalarda xatoliklar aniqlangan. Ular quyidagilar:

Yuqorida ko'rib turgan ma'lumotnomada bir qator imloviy xatoliklar yuzaga kelgan. Ma'lumotnomani beruvchi tashkilotning nomida va qarordan ilova qilingan jumlaning o'zidayoq bosh harflar imlosi, chiziqli bilan yozish, qo'shib yozish hamda harflar imlosi bilan bog'liq xatoliklar mavjud. Agar jumlaning transkripsiyaga oladigan bo'lsak, { O'zbekiston Respublikasi Vazirlar Mahkamasining 2003 yil 25 avgustdagi 365-sonli qarori bilan tasdiqlangan¹⁶³ "Nikoxlanuvchi shaxslarning tibbiy kurikdan utkazish tug'risidagi Nizomga ilova" kabi xatoliklar uchraydi. Aslida jumla quyidagi tartibda bo'lsa, yanada maqsadga muvofiq bo'lgan bo'lar edi:

{ O'zbekiston Respublikasi Vazirlar Mahkamasining 2003-yil 25-avgustdagi 365-sonli qarori bilan tasdiqlangan "Nikoxlanuvchi shaxslarning tibbiy kurikdan o'tkazish to'g'risidagi Nizomga ilova" }

Undan tashqari, tashkilot manzili nomi ostidagi qavsdagi jumlada esa harflar tushib qolishi bilan bog'liq muammo kuzatilgan. Ya'ni: { muassasi } - { muassasasi }.

Hujjat turining nomini ifodalovchi jumla ham tarkibiga imloviy xatoliklarni kiritgan:
{ SHaxslarning } - { Shaxslarning }
{ nikoxga } - { nikohga }
{ kurikdan } - { ko'rikdan }
{ utganligi } - { o'tganligi }

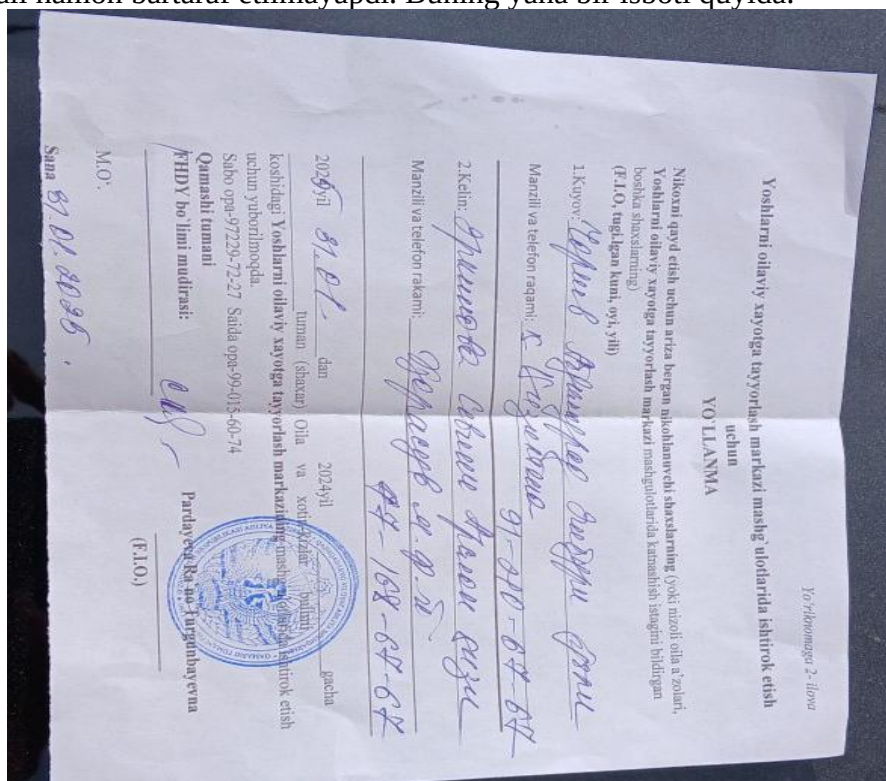
¹⁶² Azizova. D. X. "Hujjatlar va matnlarni lingvistik tahlil qilish metodikasi". - Samarqand: Samarqand Davlat Universiteti, 2017

¹⁶³ "O'zbekiston Respublikasining Oila kodeksi" 30.04.1998. 2-bo'lim. 3-bob.13-modd.

{tug'risidagi]- {to'g'risidagi}

Ma'lumotnoma matni¹⁶⁴ (u tasdiqlayotgan voqea yoki holat bayoni)da ham turli xatoliklar mavjud bo'lib, asosiy xatoliklar bo'g'iz undoshi va chuqur til orqa undoshlari hamda u va o' undoshlari bilan bog'liq xatoliklardir.

O'zbek tilining imlosi va orfografiyasini belgilovchi 1995-yil 29-dekabridagi 339-son qaroriga muvofiq "O'zbek tili imlosi" ishlab chiqilgan.¹⁶⁵ Yuqorida ko'rib chiqilgan xatoliklarni bartaraf etish chora tadbirlari qaror nizomida sanab o'tilgan. Ammo shunday bo'lsa ham bu kamchiliklar hali hamon bartaraf etilmayapdi. Buning yana bir isboti quyida:



Mazkur yo'llanmada ham yuqoridagi ma'lumotnoma bilan ayni shakldagi xatoliklar kuzatilgan. Ma'lumotnomadan farqli ravishda bunda til orqa undoshi- {k} va chuqur til orqa undoshi-{q} bilan bo'g'liq xatoliklar uchragan:

{yo'riknomaga}- {yo'riqnomaga}

{katnashish}- {qatnashish}

{boshka}- {boshqa}

{rakami}- {raqami}

{koshidagi}- {qoshidagi}

Qolaversa, xodimlarning ismi bilan bog'liq qo'pol xatolik yuzaga kelgan. Tabiiyki bu imlo bilan bog'liq xatolik emas, balki so'z qo'llashga doir uslubiy xatolikdir. Xodimlarning ism, familiyasi va sharifi ko'rsatilishi kerak bo'lgan o'rinda, huquqiy-normativ hujjatga nomuvofiq xatolik {SABO OPA} va {SAIDA OPA} tarzida keltirilmoqda. Undan tashqari tashkilot xodimlarining yozuv bilan bog'liq bo'lgan muommosini ham diqqat markazidan qochirmasligimiz lozim. Sababi mamlakatimizda 1995-yildagi 71-I-son "Lotin yozuviga asoslangan o'zbek alifbosini joriy etish to'g'risida"gi qonunidan so'ng barcha ish qog'ozlari lotin yozuviga asoslangan yangi alifboda yozilishi joriy qilingan edi.¹⁶⁶ Biroq aholining katta yoshdagi va ayrim davlat organlarida faoliyat olib boruvchi qatlami kirill alifbosiga ko'nikishgan va shu bois hujjatlar

¹⁶⁴ Aminov, M va boshqalar. "Davlat tilida ish yuritish" amaliy qo'llanma.-Toshkent. "Donishmand ziyosi", 1642022

¹⁶⁵ "O'zbek tili imlosi". 29.12.1995 339-son qaror

¹⁶⁶ "Lotin yozuviga asoslangan yangi o'zbek alifbosi". 06.05.1995. 71- I-son

tili kirill yozuvidan lotin yozuviga o'tishi qiyin kechmoqda. Bu esa, xususan, yangi turmush qurish ostonasida turgan juftliklar uchun bir qator muommolarni yuzaga keltirmoqda.

Lingvistik ekspertizaning asosiy vazifasi ham aynan yuqoridagi kabi yozuv va nutq orqali ifodalanagan ma'lumotlarning aniq va to'g'ri ifodalanishini ta'minlashdir.¹⁶⁷ Ushbu maqolada ham tor doirada bo'lsa-da, hujjatlar tili va matni hamda ularning ahamiyati bo'yicha tahlil amalga oshirildi. Chunki bunga zarurat paydo bo'lib bormoqda. Bu zarurat hujjatlarning aniq va tushunarli bo'lishini ta'minlash, shuningdek huquqiy muommolarga yo'l qo'yimaslik uchun lozimdir. Bu o'z navbatida fuqarolar uchun tushunarli va aniq huquqiy normalarni ishlab chiqish imkonini yaratadi.

FHDYO hujjatlari loyihalarini lingvistik ekspertizadan o'tkazish, nafaqat ularning tildagi ifoda aniqligini, balki huquqiy, uslubiy va kommunikativ mezonlarga muvofiqligini ta'minlashda muhim vosita hisoblanadi. Ushbu maqolada hujjatlar matnining aniqligi, tushunarligi va rasmiy-uslubiy me'yorlarga mosligi ekspertiza jarayonida qanday baholanishi lozimligi yoritildi. Shuningdek, hujjatlarda uchrashi mumkin bo'lgan tipik til xatolari, ularning sabablari va oldini olish yo'llari tahlil qilindi.

Lingvistik ekspertiza huquqiy amaliyotda ishonchlilikni oshiradi, inson omili ta'sirini kamaytiradi va notog'ri talqinlar ehtimolini minimallashtiradi. Ayniqsa, FHDYO tizimida qo'llaniladigan ariza, yo'llanma, qaror, ma'lumotnoma kabi hujjatlarning mazmunan va shaklan to'g'ri tuzilishi fuqarolarning huquq va manfaatlarini himoya qilishda muhim ahamiyatga ega. Shu boisdan lingvistik ekspertiza metodologiyasini takomillashtirish, amaliyotda samarali qo'llash va mutaxassislar malakasini oshirish bugungi kunda dolzarb vazifalardan biridir. Kelgusida huquqiy-normativ hujjatlar loyihalarini avtomatlashtirilgan lingvistik tahlil tizimlari orqali baholash, ularni aniq me'yorlar asosida standartlashtirish istiqbollari mavjud bo'lib, bu yo'nalishda izchil ilmiy izlanishlar olib borilishi zarur.

Foydalanilgan adabiyotlar:

1. "O'zbek tili imlosi". 29.12.1995 339-son qaror
2. "Lotin yozuviga asoslangan yangi o'zbek alifbosi". 06.05.1995. 71- I-son
3. "O'zbekiston Respublikasining Oila kodeksi" 30.04.1998. 2-bo'lim. 3-bob.13-modda.
4. "O'zbekiston Respublikasining Oila kodeksi" 30.04.1998. 2-bo'lim. 3-bob.17-modda.
5. Aminov. M va boshqalar. "Davlat tilida ish yuritish" amaliy qo'llanma.-Toshkent. "Donishmand ziyosi", 1682022
6. Azizova. D. X. "Hujjatlar va matnlarni lingvistik tahlil qilish metodikasi".-Samarqand: Samarqand Davlat Universiteti, 2017
7. Mamatov. R. S. "Lingvistik ekspertizaning huquqiy va amaliy aspektlari"-Toshkent. "Yuridik ta'lim nashri", 2022.

¹⁶⁷ Mamatov. R. S. "Lingvistik ekspertizaning huquqiy va amaliy aspektlari"-Toshkent. "Yuridik ta'lim nashri", 2022.

QORAQALPOQ TILINI KORPUS LINGVISTIKASI ASOSIDA TADQIQ QILISH

Xudaybergenova Gozzal Kenesbay qizi

Qoraqalpoq davlat universiteti

Kompyuter lingvistikasi ta'lim yo'nalishi 2-kurs talabasi

E-mail: xudaybergenovagozzal670@gmail.com

Annotatsiya: Bu maqola Qoraqalpoq tili korpusi va uning lingvistik tahlili haqida. Unda korpusni yaratish, saqlash va tahlil qilish texnologiyalari, jumladan, morfologik analizatorlar, tokenizatsiya va lemmatizatsiya kabi jarayonlar muhokama qilinadi. Shuningdek, maqola qoraqalpoq tilining raqamli maydondagi rivojlanishi va global kommunikatsiya sohasidagi ahamiyatini ko'rsatadi.

Kalit so'zlar: Qoraqalpoq tili, korpus, morfologik tahlil, tokenizatsiya, lemmatizatsiya, lingvistika, raqamli til texnologiyalari, global kommunikatsiya.

Аннотация: Эта статья о корпусе каракалпакского языка и его лингвистическом анализе. В нем обсуждаются технологии создания, хранения и анализа корпуса, включая такие процессы, как морфологические анализаторы, токенизация и лемматизация. Также в статье показано развитие каракалпакского языка в цифровом пространстве и его значение в сфере глобальной коммуникации.

Ключевые слова: Каракалпакский язык, корпус, морфологический анализ, токенизация, лемматизация, лингвистика, цифровые языковые технологии, глобальная коммуникация.

Abstract: This article is about the Karakalpak language corpus and its linguistic analysis. It discusses corpus creation, storage, and analysis technologies, including processes such as morphological analyzers, tokenization, and lemmatization. The article also shows the development of the Karakalpak language in the digital space and its importance in the field of global communication.

Keywords: Karakalpak language, corpus, morphological analysis, tokenization, lemmatization, linguistics, digital language technologies, global communication.

Qoraqalpoq tili - Turkiy tillar oilasining qipchoq guruhiga kirib, Markaziy Osiyoda, asosan Qoraqalpog'iston Respublikasida va qo'shni davlatlarda yashovchi qoraqalpoq xalqi tomonidan so'zlashadi.

Qoraqalpoq tili haqida. Qoraqalpoq xalqi va ularning tilining shakllanishi uzoq tarixiy jarayonlarga borib taqaladi. XIII-XIV asrlarda qoraqalpoq qabilalari turli hududlarda yashab, asta-sekin o'ziga xos til va madaniyatni rivojlantirgan. Qoraqalpoq tili o'z rivojlanishida qadimgi qipchoq va boshqa turkiy tillardan ta'sir olgan. Tarixiy yozma yodgorliklar qoraqalpoq tilining qadimiy ildizlariga ega ekanligini ko'rsatadi [O'teganov S., 1985]. Nemets tilshunosi I.X. Adelung "Mitridata" asarida Orenburg tatarlari va Turkiston tatarlari tillarining biri sifatida qoraqalpoq tilini ikki joyda tilga oladi va Orenburg guruhida boshqird, ufa tatarlari va mesherlar bilan tili jihatidan yaqin ekanligini ko'rsatadi [Sh. Abdunazimov, 2008, 7].

Qoraqalpoq tilining grammatikasi agglyutinatив (ya'ni, so'zlarga qo'shimchalar qo'shish orqali grammatik ma'no ifodalash) tizimga asoslanadi.

Asosiy grammatik xususiyatlari:

- ✓ So'z tartibi: Sodda gaplarda "subyekt + obyekt + predikat" (SOV) tartibi ustunlik qiladi.
- ✓ Fe'l zamonlari: Hozirgi, o'tgan va kelasi zamon shakllari bor.
- ✓ Otlarning kelishik, son va egalik kategoriyalari rivojlangan.
- ✓ Sifat, ravish va olmoshlar mustaqil so'z turkumlari sifatida faol qo'llanadi.

Bugungi kunda qoraqalpoq tili Qoraqalpog'iston Respublikasining davlat tillaridan biri sifatida rasmiy maqomga ega. Maktablarda, universitetlarda qoraqalpoq tilida ta'lim beriladi, gazeta-jurnallar, kitoblar, ilmiy va badiiy asarlar qoraqalpoq tilida chop etilmoqda. Shunga qaramay, zamonaviy axborot texnologiyalari va globallashuv oqibatida tilning rivojlanishi va saqlanib qolishi uchun alohida e'tibor talab etilmoqda. Xususan, qoraqalpoq tilining raqamli muhitdagi imkoniyatlarini kengaytirish, korpus lingvistikasi asosida tadqiq etish.

Bugungi kunda tilshunoslik sohasida zamonaviy texnologiyalar va katta hajmdagi matnlar bilan ishlash imkoniyati tilni o'rganish usullarini tubdan o'zgartirib yubordi. Ana shunday zamonaviy yo'nalishlardan biri bu - korpus lingvistikasidir.

Korpus lingvistikasi - tilni real hayotda qo'llanilgan matnlar asosida o'rganadigan fan sohasi. Bu yo'nalish til hodisalarini empirik asosda, ya'ni ko'plab haqiqiy matnlar asosida tekshirish imkonini beradi. McEnery va Wilson ta'kidlaganidek: "Korpus lingvistikasi - tilni real foydalanishdagi matnlar asosida o'rganadigan zamonaviy tarmoq." 2-26]. Bu metodologiya, bir tomondan, nazariy gipotezalarni sinovdan o'tkazish uchun kerakli materiallar beradi, ikkinchi tomondan, yangi nazariyalar ishlab chiqishga ham xizmat qiladi. Beyker fikricha: "Korpuslar tadqiqotchilarga tilni sun'iy emas, balki tabiiy sharoitdagi holatida o'rganish imkonini beradi." [Baker P., 2006; 33-bet. Bu esa tilshunos olimlarga nafaqat individual misollarni, balki til hodisalarining umumiy tendensiyalarini aniqlash imkonini yaratadi. Bugungi globallashuv davrida har bir milliy til uchun raqamli korpus yaratish dolzarb ahamiyat kasb etmoqda. O'tayev ta'kidlaydi: "Bugungi globallashuv davrida har bir til uchun raqamli matn korpuslarini yaratish uning rivojlanishida muhim rol o'ynaydi."

Korpus lingvistikasi nima va u nima uchun dolzarb. So'nggi yillarda tilshunoslikda matn va nutq materiallarini keng ko'lamda o'rganish imkoniyatlarini ta'minlaydigan yangi yondashuvlar shakllandi. Ulardan eng muhimlaridan biri - korpus lingvistikasidir. Korpus lingvistikasi real hayotda yaratilgan matnlar asosida tilni chuqur, tizimli o'rganishga imkon beradigan fan sohasi hisoblanadi. Bu yondashuv nazariy fikrlarga emas, balki empirik ma'lumotlarga, ya'ni haqiqiy nutq namunalari asosida o'rganishga asoslanganligi bilan ajralib turadi. Korpus lingvistikasi - bu real foydalanilgan matnlarni tizimli ravishda to'plab, ularni tahlil qilish orqali tilning haqiqiy qo'llanilish xususiyatlarini o'rganishga asoslangan yondashuvdir [McEnery T., Wilson A., 1996; 2].

Korpus lingvistikasi klassik lingvistika usullaridan farqli ravishda ilmiy tadqiqotlarda raqamli dalillarni asos sifatida namoyon etadi. U grammatik hodisalarining, leksik birliklarning, uslubiy vositalarning real qo'llanilish darajasini aniqlashda, til me'yorlarini tasdiqlash yoki rad etishda juda qulay va samarali metod hisoblanadi. Bugungi raqamli texnologiyalar davrida korpuslar katta ma'lumotlar (Big Data) konsepsiyasining bir qismi sifatida tilshunos olimlarga aniq va keng qamrovli tahlil imkoniyatlarini yaratib bermoqda. Natijada, til hodisalarining sonli tavsifi, ularning o'zgarish dinamikasini o'rganish va til rivojlanishining real yo'nalishlarini aniqlash imkoniyati sezilarli darajada ortdi. Korpus lingvistikasi nafaqat tilshunoslikda, balki tarjima nazariyasi, sun'iy intellekt, ma'lumotlarni qayta ishlash, avtomatik tarjima tizimlari kabi ko'plab sohalarida ham keng qo'llanilmoqda. Ayniqsa, kichik va o'ziga xos tillar uchun korpus yaratish va ular asosida ilmiy tahlil o'tkazish - milliy madaniyatni saqlash va rivojlantirish uchun muhim hisoblanadi. Hozirgi globallashuv jarayonida har bir milliy til uchun korpuslar yaratish zarurati yanada dolzarb ahamiyat kasb etdi. Tilni empirik ma'lumotlarga asoslanib o'rganish, nazariy mulohazalarga qaraganda ko'proq aniqlik va ishonchlikni ta'minlaydi.

Qoraqalpoq tilini korpus asosida o'rganish. Hozirgi zamonda milliy tillarni rivojlantirish va ularning raqamli makonda o'z o'rnini mustahkamlash uchun zamonaviy ilmiy-uslubiy yondashuvlarga talab kuchaymoqda. Ayniqsa, kichik tilda so'zlashuvchi jamoalar uchun bu jarayon juda muhim ahamiyatga ega. Qoraqalpoq tili ham shunday tillardan biri.

Qoraqalpoq tilining o'ziga xos fonetik, grammatik va leksik xususiyatlari bo'lib, ularni to'liq o'rganish va hujjatlashtirish uchun korpus lingvistikasi asosida tadqiqotlar olib borish bugungi kunda zarur hisoblanadi. Real matnlar asosida tuzilgan til korpuslari qoraqalpoq tilining amaliy

qo'llanilishini aniq ko'rsatishga, grammatik me'yorlar va tendensiyalarni aniqlashga imkoniyat yaratadi. Korpus asosida tilni o'rganish kichik tillarning o'zini oqlashiga, ularning zamonaviy ilm olamida o'z o'rnini topishiga xizmat qiladi [Nazarov G., Xudoyberganov A, 2020; 45].

Qoraqalpoq tili korpusini yaratish bir necha jihatdan ahamiyatli: Birinchidan, tilni tizimli ravishda tahlil qilish va real nutqdagi hodisalarni aniqlash imkoniyati kengayadi. Ikkinchidan, qoraqalpoq tilining saqlanishi va rivojlanishiga xizmat qiladigan mustahkam raqamli resurs yaratiladi. Uchinchidan, yosh avlod uchun qoraqalpoq tilida zamonaviy ta'lim, tarjima va tahlil dasturlarini ishlab chiqish imkoniyati paydo bo'ladi.

Bugungi globallashtirish va raqamli kommunikatsiyalar asrida o'z ona tilining raqamli ko'rinishda mavjudligi - milliy madaniyat va tilni saqlash uchun muhim vositadir. Korpusga asoslangan tadqiqotlar qoraqalpoq tilining lug'at boyligini, grammatik tuzilishini, og'zaki va yozma tildagi farqlarni tahlil qilishda ishonchli asos bo'lib xizmat qiladi.

Qoraqalpoq tilida korpus yaratuvchanlik imkoniyatlari. Qoraqalpoq tilida zamonaviy korpus yaratish - ilmiy va amaliy faoliyatda keng imkoniyatlar ochadigan muhim yo'nalishlardan biri. Bugungi kunga qadar qoraqalpoq tili uchun to'liq va tizimli raqamli korpuslar yetarli darajada shakllanmagan. Shuning uchun, korpus tuzish ishlarini boshlash zamonaviy tilshunoslik uchun zaruratga aylangan. Korpus yaratish uchun qoraqalpoq tilining xususiyatlarini hisobga olgan holda quyidagi yo'nalishlarda ishlash zarur:

✓ Matn tanlash va to'plash: Adabiy, ilmiy, ommaviy axborot vositalari, internet resurslaridan qoraqalpoq tilidagi sifatli va tabiiy matnlar to'planadi.

✓ Matnlarni belgilash: Morfologik, sintaktik va semantik belgilar qo'yiladi. Bu tilni chuqurroq tahlil qilish uchun asos bo'ladi.

✓ Meta ma'lumotlar qo'shish: Har bir matn uchun muallif, janr, yozilgan yili, matn turi kabi ma'lumotlar tayyorlanadi.

✓ Morfologik va sintaktik tahlil: So'z shakllari tahlili, gap qurilmalari va ularning tuzilishini aniqlash ishlari olib boriladi.

Korpus lingvistikasi uchun zarur materialni to'plash va belgilash jarayoni aniq va tizimli me'yorlarga asoslangan bo'lishi zarur.[Leech G., 1992; 14].

Qoraqalpoq tilida korpus yaratishda ayniqsa quyidagilarni e'tiborga olish muhim: Dialektik farqlarni hisobga olish, ya'ni qoraqalpoq tilining har xil sheva va lahjalaridan namunalar to'plash. Tarixiy va zamonaviy matnlarni uyg'un tarzda tanlash. Internet texnologiyalaridan foydalangan holda korpusni muntazam yangilab borish.

Xususan, birinchi navbatda qoraqalpoq tilidagi badiiy adabiyotlar, ilmiy maqolalar, ommaviy axborot vositalari, darsliklar va blog kabi zamonaviy matnlar asosida korpus yaratish mumkin. Bu nafaqat lingvistik tadqiqotlar uchun, balki qoraqalpoq tilida avtomatik tarjima, grammatik tekshirish, so'z tahlil kabi amaliy dasturlarni ishlab chiqishda ham asosiy resurs bo'ladi.

Hozirgi vaqtda qoraqalpoq tiliga oid to'liq, keng qamrovli elektron korpus yo'q. Hozirgi vaqtda faqat ayrim kichik loyihalar doirasida qoraqalpoq matnlarini raqamlashtirish va ularni elektron bazaga joylashtirish bo'yicha harakatlar amalga oshirilgan. Misol uchun, O'zbekiston Fanlar akademiyasi Qoraqalpog'iston bo'limi tomonidan ayrim qoraqalpoq adabiy matnlari va ilmiy maqolalari elektron shaklga o'tkazilgan. Ammo bu materiallar hali to'liq morfologik, sintaktik va semantik tahlil qilinmagan. Xorijiy manbalar ham qoraqalpoq tili uchun alohida korpus yaratishni ilgari surmagan. Shu bois qoraqalpoq tili korpusini yaratish tashabbusi o'zbek va qoraqalpoq tilshunolariga tushmoqda. Ayrim milliy tillar uchun mustaqil korpus yaratish - ularning ilmiy va amaliy rivojlanishi uchun zarur shartdir.[Romer, 2008; 77].

So'z shakllarini aniqlash, grammatik kategoriyalarga ajratib belgilash ishlari amalga oshiriladi. Bu matnlarning struktur tahliliga, grammatik hodisalarni tizimli tarzda o'rganishga imkoniyat yaratadi.

✓ Korpus bazasini yaratish va boshqarish: To‘plangan va belgilangan matnlar maxsus bazada saqlanadi. Ushbu baza foydalanuvchilar uchun qulay qidiruv va tahlil vositalarini taqdim etishi kerak.

✓ Korpusga qo‘llaniladigan texnologiyalar: Avtomatik morfologik analizatorlar, matnlarni tokenizatsiyalash, lemmatizatsiyalash va parserlash tizimlari qo‘llaniladi. Korpus qayta ishlashda Python, NLTK, SpaCy kabi zamonaviy dasturiy vositalar ishlab chiqiladi.

Qoraqalpoq tili korpusining kutilayotgan foydalari: til materiallarini tizimli ravishda saqlash va o‘rganish, grammatik me‘yorlarni aniqlash, yozish va gapirish standartlarini belgilash, avtomatik tarjima, sun‘iy intellekt va ma‘lumotlarni qayta ishlash dasturlariga asos bo‘lish, til o‘qitish, adabiyotshunoslik va lingvistik tadqiqotlarda qo‘llaniladigan natijalarni yaratish, qoraqalpoq tilining raqamli maydondagi mavqeini kuchaytirib, global kommunikatsiya sohasidagi o‘rnini mustahkamlashga yordam beradi.

Xulosa qilib aytganda, qoraqalpoq tili korpus lingvistikasi bo‘yicha tadqiqotlar tilimizning rivojlanishi va saqlanishi uchun strategik ahamiyatga ega. Zamonaviy texnologiyalar yordamida tuzilgan korpuslar orqali Qoraqalpoq tilining fonetik, grammatik va leksik tarkibi chuqur va samarali o‘rganiladi, bu esa tilimizning kelajakdagi rivojiga mustahkam poydevor bo‘lib xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Sh. Abdinazimov “Qoraqalpoq tili tarixi”, 2008.Nókis.
2. Ótegenov S. Qoraqalpoq tiliniń tarixı. – Nukus: Qaraqalpaqstan, 1985.
3. McEnery T., Wilson A., Corpus Linguistics: An Introduction, 1996.
4. Baker P., Using Corpora in Discourse Analysis, 2006.
5. Óteev N., Zamanagóy lingvistika hám sanlı texnologiyalar, 2021.
6. Nazarov G., Xudoyberganov A., Tilshunoslik va raqamli texnologiyalar, 2020.
7. Leech G., Corpus Linguistics: Method, Theory and Practice, 1992.
8. Römer, Corpus Research Applications, 2008

DUNYO KOMPYUTER LINGVISTIKASIDA SOHAVIY TEZAVURUS YARATISHGA OID TAJRIBALAR TAHLILI

Ubaydullayeva Ziroatxon

O'zbekiston Milliy universiteti

Kompyuter lingvistikasi yo'nalishi 2-kurs magistranti

Annotatsiya: Maqolada dunyo kompyuter lingvistikasida sohaviy tezavrus yaratish bo'yicha olib borilgan tadqiqotlar tahlil qilinadi. Soha bo'yicha yetakchi ilmiy yondashuvlar, modellash metodlari va tezavrus tuzish jarayonida qo'llanilayotgan texnologiyalar ko'rib chiqiladi. Shuningdek, turli tillardagi sohaviy tezavruslarning afzalliklari va kamchiliklari muhokama qilinadi hamda ularning tabiiy tilni qayta ishlash (NLP) tizimlaridagi roli yoritiladi. Tadqiqot natijalari sohaviy tezavruslarni yaratishda samarali yondashuvlarni aniqlash va ularning amaliy qo'llanilish imkoniyatlarini kengaytirishga yo'naltirilgan.

Kalit so'zlar: Kompyuter lingvistikasi, sohaviy tezavrus, tabiiy tilni qayta ishlash (NLP), bilim bazasi, semantik tahlil, leksikografiya, ma'lumotlarni tuzish, ontologiya.

Texnologiyalar rivojlanishi bilan inson va kompyuter o'rtasidagi muloqot yanada chuqurlashib bormoqda. Sun'iy intellekt (AI) va tabiiy tilni qayta ishlash (NLP) texnologiyalari kundalik hayotimizning ajralmas qismiga aylanmoqda. Ushbu jarayonda tilshunoslik va informatika sohalarining kesishish nuqtasi sifatida tezavrus tushunchasi alohida ahamiyat kasb etadi. Xususan, sohaviy tezavruslar muayyan ilmiy yoki amaliy sohaga oid terminlarni tizimlashtirish, ularning o'zaro bog'liqligini belgilash va axborot izlash jarayonlarini soddalashtirishga xizmat qiladi.

Dunyo bo'ylab ilmiy va texnologik rivojlanishning ortib borishi natijasida turli tillarda sohaviy tezavruslarni yaratish bo'yicha keng ko'lamli tadqiqotlar olib borilmoqda. Ayrim mamlakatlar ushbu yo'nalishda oldinga siljib, o'zlarining yirik loyihalarini ishga tushirgan. Masalan, AQShning UMLS (Unified Medical Language System) tibbiyot sohasida qo'llaniladigan eng keng qamrovli terminologik tizimlardan biri bo'lsa, Yevropada EuroVoc huquq va siyosiy fanlarga oid terminlarni tartibga solish uchun yaratilgan. Shuningdek, Osiyo davlatlari ham NLP sohasida o'z hissasini qo'shmoqda. Masalan, Xitoyning HowNet tizimi va Yaponiyaning EDR Electronic Dictionary resurslari so'z ma'nolari va ularning kontekstual bog'liqligini o'rganishda muhim ahamiyatga ega.

Sohaviy tezavruslarni yaratish usullari har xil bo'lishi mumkin. Ba'zi hollarda ularni tilshunos va mutaxassislar qo'lda ishlab chiqadi, bu esa aniqlik va sifat jihatidan yuqori natija beradi. Biroq bunday yondashuv ko'p vaqt va resurs talab qiladi. Shu sababli, bugungi kunda mashinali o'rganish va neyron tarmoqlar asosida avtomatik tezavrus yaratish usullari keng ommalashmoqda. Word2Vec, GloVe kabi modellar yordamida matnlardagi so'zlar semantik jihatdan tahlil qilinadi va ularning o'zaro bog'liqligi aniqlanadi.

Tezavruslar nafaqat lingvistik tadqiqotlar uchun, balki real hayotda turli sohalarida ham keng qo'llaniladi. Masalan, tibbiyotda UMLS tizimi shifokor va tadqiqotchilarga kasallik nomlari, dori vositalari va davolash usullari o'rtasidagi bog'liqlikni tushunishda yordam beradi. Yuridik sohada esa EuroVoc qonunchilik hujjatlarini indeksatsiya qilish va huquqiy axborotni tezkor izlash imkonini beradi. Muhandislik va texnologiya sohalarida esa tezavruslar hujjatlarni avtomatik tahlil qilish va texnik atamalarni standartlashtirishga xizmat qiladi.

Shunday bo'lsa-da, sohaviy tezavruslar yaratishda qator muammolar ham uchraydi. Eng katta muammo – til va madaniy tafovutlar. Har bir til va madaniyatda so'zlarning o'ziga xos ma'nolari va qo'llanilish konteksti bor. Shu sababli, ko'p tilli tezavruslarni yaratish va ularni global miqyosda qo'llash qiyin vazifa hisoblanadi. Yana bir muhim muammo – ma'lumotlar bazalarining dinamikasi. Terminologiyalar doimiy ravishda o'zgarib, yangilanib boradi, bu esa tezavruslarni muntazam ravishda yangilab borishni talab qiladi.

Kelajakda sohaviy tezauruslar yanada rivojlanishi va NLP, sun'iy intellekt tizimlari bilan integratsiyalashishi kutilmoqda. Masalan, ChatGPT, Google Bard, DeepL kabi ilg'or sun'iy intellekt modellariga tezaurus ma'lumotlari kiritilib, yanada aniq va semantik jihatdan boy javoblar olish imkoniyati paydo bo'ladi. Bundan tashqari, blokcheyn texnologiyalari va kognitiv hisoblash tizimlari yordamida tezauruslarni avtomatik ravishda yangilash mexanizmlari ishlab chiqilishi mumkin. Sohaviy tezauruslar kompyuter lingvistikasi va tabiiy tilni qayta ishlash tizimlarining ajralmas qismiga aylanib bormoqda. Ular axborot qidirish, matnlarni tahlil qilish va sun'iy intellekt tizimlarini rivojlantirishda muhim rol o'ynaydi. Kelajakda ushbu tizimlar yanada mukammallashtirilsa, ilm-fan va amaliy sohalarida katta yutuqlarga erishish mumkin bo'ladi. Bundan tashqari, sohaviy tezauruslarning kelajakdagi rivojlanishi bir necha yo'nalishda jadallashishi kutilmoqda. Avtomatik tezaurus yaratish va yangilash jarayoni sun'iy intellekt yordamida yanada takomillashib, inson ishtirokisiz ham yuqori sifatli ma'lumotlar bazasini shakllantirish imkoniyatini yaratadi. Hozirgi paytda aksariyat tezauruslar qo'lda ishlab chiqilgan bo'lib, bu katta vaqt va resurs talab qiladi. Biroq, GPT kabi til modellarining rivojlanishi bilan bu jarayonni avtomatlashtirish imkoni paydo bo'lmoqda. Masalan, OpenAI, Google DeepMind va Meta AI kabi tashkilotlar tomonidan ishlab chiqilgan ilg'or NLP tizimlari so'zlarning semantik aloqalarini aniqlash va tasniflash bo'yicha katta natijalarga erishmoqda.

Kelajakda blokcheyn va kognitiv hisoblash texnologiyalarining integratsiyasi sohaviy tezauruslarni yanada ishonchli va shaffof qilishga yordam berishi mumkin. Blokcheyn orqali tezaurus ma'lumotlari saqlansa, ularning haqiqiyligini tekshirish, so'zlarning o'zgarish tarixini kuzatish va noto'g'ri ma'lumotlarning kiritilishining oldini olish osonlashadi. Shu bilan birga, bulut texnologiyalari yordamida tezaurus ma'lumotlari real vaqt rejimida yangilanib, barcha foydalanuvchilar uchun qulay va tezkor foydalanish imkoniyati yaratiladi. Shuningdek, ko'p tilli va ko'p madaniyatli tezauruslar yaratish ehtiyoji ham ortib bormoqda. Bugungi kunda aksariyat tezauruslar bitta yoki ikkita tilga mo'ljallangan bo'lsa, kelajakda sun'iy intellekt yordamida global miqyosda barcha tillar va lahjalarga moslashtirilgan tezauruslar ishlab chiqilishi mumkin. Bu esa turli tillarda ilmiy izlanishlar olib borayotgan tadqiqotchilar, tarjimonlar va texnologik kompaniyalar uchun katta foyda keltiradi.

Bundan tashqari, sohaviy tezauruslarning sun'iy intellekt bilan integratsiyasi turli sohalaridagi texnologik rivojlanishga katta ta'sir o'tkazishi mumkin. Masalan, avtomatlashtirilgan tibbiy diagnostika tizimlari, huquqiy hujjatlarni tahlil qiluvchi yuridik sun'iy intellekt tizimlari yoki ilmiy tadqiqotlarni tezlashtirishga yordam beruvchi aqlli ilmiy ma'lumotlar bazalari sohaviy tezauruslar bilan bog'langan holda ishlasa, natijalar ancha samarali bo'ladi. Shu bilan birga, tezauruslar asosida yaratilgan sun'iy intellektli virtual yordamchilar odamlarning kundalik hayotida muhim rol o'ynashi mumkin. Masalan, Google Assistant, Siri yoki Alexa kabi ovoqli yordamchilar sohaviy tezaurus ma'lumotlaridan foydalansa, murakkab texnik yoki ilmiy tushunchalarni ham to'g'ri izohlash va tushuntirish imkoniga ega bo'ladi.

Kelajakdagi eng katta muammolardan biri esa axborot xavfsizligi va shaxsiy ma'lumotlarning himoyasi bo'lib qolmoqda. Tezauruslar tobora ko'proq sohalariga integratsiyalashgani sari, ularga kiritiladigan ma'lumotlarning ishonchiligi va maxfiyligi muhim masala bo'lib qoladi. Shuning uchun ham sun'iy intellekt asosida ishlovchi tezauruslar axborot filtrlash, noto'g'ri yoki yolg'on ma'lumotlarni aniqlash va yo'q qilish kabi funksiyalarga ega bo'lishi lozim.

Sohaviy tezauruslarning ahamiyati va ularning kelajakdagi rivojlanish istiqbollari haqida batafsil o'ylab ko'rsak, bu yo'nalishning nafaqat ilm-fan va texnologiyada, balki kundalik hayotimizda ham katta o'zgarishlarga sabab bo'lishi aniq.

Bugungi kunda raqamli transformatsiya barcha sohalariga o'z ta'sirini ko'rsatmoqda. Turli tarmoqlarda, xoh u tibbiyot bo'lsin, xoh huquq, biznes yoki ta'lim, axborotning tezkor va samarali qidirilishi muhim omilga aylangan. Sohaviy tezauruslar esa bu jarayonda axborotni to'g'ri tasniflash, mantiqiy bog'liqliklarni aniqlash va insonning qidiruv natijalaridan maksimal foyda olishini ta'minlash orqali yordam beradi. Sun'iy intellekt va mashinali o'rganish algoritmlarining

rivojlanishi bilan sohaviy tezauruslarning roli yanada kuchayib bormoqda, chunki ular turli ma'lumotlar bazalarini o'zaro bog'lash va ma'lumotlarni tezkor tahlil qilish imkonini beradi. Misol uchun, tibbiyot sohasida sohaviy tezauruslar shifokorlarga va tibbiy tadqiqotchilarga kasallik belgilari, dorilar va muolajalar bo'yicha keng qamrovli ma'lumotlarni taqdim etishda katta yordam beradi. Bu esa kasalliklarni erta aniqlash va aniq tashxis qo'yish jarayonini tezlashtirish imkonini yaratadi. Masalan, UMLS (Unified Medical Language System) tizimi dunyodagi eng yirik tibbiy terminologik resurslardan biri bo'lib, turli tillardagi ma'lumotlarni bir joyda jamlaydi. Shuningdek, huquq sohasida sohaviy tezauruslar sud hujjatlari, qonunlar va normativ hujjatlarni tartibga solishda katta ahamiyat kasb etadi. Yuridik axborotni tezkor izlash va huquqiy masalalarni hal qilishda EuroVoc kabi tizimlar qo'llaniladi. Kelajakda sun'iy intellekt yordami bilan yuridik maslahat xizmatlarini avtomatlashtirish va sud hujjatlarini tahlil qilish kabi jarayonlar yanada rivojlanishi mumkin.

Bundan tashqari, biznes va iqtisodiyotda sohaviy tezauruslardan foydalanish bozor tahlili, moliyaviy prognoz qilish va xatarlarni baholashda muhim ahamiyat kasb etmoqda. Masalan, big data texnologiyalari yordamida kompaniyalar o'z mijozlari haqida yanada batafsil ma'lumot to'plashlari va ularning ehtiyojlarini aniqlashlari mumkin. Bu jarayonda sohaviy tezauruslar katta hajmdagi ma'lumotlarni tizimlashtirish va maqsadli tahlil qilish imkonini beradi.

Ta'lim va ilmiy tadqiqotlar sohasida sohaviy tezauruslarning rivojlanishi ilmiy ma'lumotlar bazalarini yanada samarali tashkil qilishga yordam beradi. Hozirda tadqiqotchilar Google Scholar, Scopus yoki Web of Science kabi platformalardan foydalanishadi. Kelajakda esa aqlli ilmiy yordamchilar (AI-based research assistants) yaratilishi mumkin bo'lib, ular tadqiqotchilarga eng dolzarb ilmiy maqolalarni tavsiya qilish, terminlarni tushuntirish va ilmiy izlanishlarni tezlashtirishda yordam beradi.

Sohaviy tezauruslar va tabiiy tilni qayta ishlash (NLP) texnologiyalarining integratsiyasi esa inson-kompyuter o'zaro ta'sirini yangi bosqichga olib chiqmoqda. Bugungi kunda chatbotlar va virtual yordamchilar foydalanuvchilarga turli sohalar bo'yicha maslahat berish, hujjatlarni avtomatik ravishda yaratish va axborotni tizimlashtirish imkoniyatini taqdim etmoqda. Kelajakda bunday tizimlar yanada intellektual bo'lib, turli sohalar uchun moslashtirilgan maxsus sohaviy chatbotlar yaratish imkoniyati yuzaga keladi. Shu bilan birga, sohaviy tezauruslarning ko'p tilli modelga asoslangan holda rivojlanishi global muloqotni yangi bosqichga olib chiqishi mumkin. Hozirgi kunda mashinalar tomonidan tarjima qilish texnologiyalari, masalan, Google Translate yoki DeepL, ancha rivojlangan bo'lsa-da, hali ham mukammal emas. Ko'p tilli sohaviy tezauruslar ushbu kamchiliklarni bartaraf etishda muhim rol o'ynashi mumkin. Masalan, bitta tilda ifodalangan terminologiyani boshqa tillarda ham to'g'ri kontekstda tarjima qilish imkoniyati paydo bo'ladi.

Biroq, sohaviy tezauruslarni yaratish va ulardan foydalanishda hali ham bir qancha muammolar mavjud. Eng asosiy muammolardan biri – til va madaniy tafovutlar. Har bir til o'ziga xos lug'aviy va semantik xususiyatlarga ega bo'lib, bu esa universal tezaurus yaratishni qiyinlashtiradi. Shuningdek, mavjud tezauruslarning dinamikasi va doimiy ravishda yangilanish zarurati ham muhim masala hisoblanadi. Chunki fan va texnologiya rivojlanishi bilan yangi atamalar va tushunchalar paydo bo'lib, ularni o'z vaqtida tezauruslarga qo'shish lozim. Kelajakda ushbu muammolarni hal qilish uchun avtomatlashtirilgan va sun'iy intellektga asoslangan tezaurus yaratish metodlari yanada rivojlanishi mumkin. Mashinali o'rganish texnologiyalarining takomillashishi natijasida yangi tushunchalarni avtomatik ravishda aniqlash va ularni tegishli kategoriya va kontekstga mos ravishda joylashtirish imkoniyati oshadi.

Xulosa qilib aytganda, sohaviy tezauruslarning roli tobora ortib bormoqda va kelajakda turli sohalar uchun maxsus moslashtirilgan aqlli tizimlar shaklida yanada rivojlanishi kutilmoqda. Bunday tizimlarning rivojlanishi ilm-fan, ta'lim, tibbiyot, huquq va biznes sohalarida katta inqilobga sabab bo'lishi mumkin. Sun'iy intellekt va tabiiy tilni qayta ishlash texnologiyalari bilan bog'liq holda, sohaviy tezauruslar nafaqat ma'lumotlarni tizimlashtirish, balki axborotga tezkor va to'g'ri kirish imkoniyatini ham ta'minlaydi.

Kelajakda blokcheyn, bulut texnologiyalari va kognitiv hisoblash asosida ishlab chiqilgan tezaurus tizimlari yanada ishonchli, samarali va intellektual bo'lishi kutilmoqda. Bunday taraqqiyot esa ilmiy tadqiqotlar, texnologik innovatsiyalar va sun'iy intellektga asoslangan xizmatlarni yangi bosqichga olib chiqadi. Shunday qilib, dunyo bo'ylab kompyuter lingvistikasi sohasida sohaviy tezauruslar yaratish bo'yicha olib borilayotgan tadqiqotlar yangi texnologiyalar bilan uyg'unlashgan holda davom etmoqda. Kelajakda ushbu tizimlar yanada mukammal bo'lib, turli sohalaridagi muammolarni hal qilishda muhim vosita bo'lib xizmat qilishi aniq. Axborot texnologiyalarining uzluksiz rivojlanishi bilan inson va sun'iy intellekt o'rtasidagi aloqalar chuqurlashib, bilimlarni samarali boshqarish imkoniyati yanada kengayadi.

Xulosa qilib aytganda, sohaviy tezauruslar tabiiy tilni qayta ishlash, sun'iy intellekt va ma'lumotlarni boshqarish sohalarining ajralmas qismiga aylanmoqda. Ularning rivojlanishi ilmiy va texnologik taraqqiyotga katta hissa qo'shishi shubhasiz. Kelajakda avtomatlashtirilgan, ko'p tilli va intellektual tezaurus tizimlarining paydo bo'lishi inson va kompyuter o'rtasidagi muloqotni yanada mukammal qilish bilan birga, global axborot makonini rivojlantirishga ham katta ta'sir ko'rsatadi. Shunday qilib, dunyo kompyuter lingvistikasi olamida sohaviy tezauruslar ustida olib borilayotgan tadqiqotlar yangi texnologiyalar bilan uyg'unlashgan holda davom etmoqda. Kelajakda ushbu tizimlar yanada mukammal bo'lib, turli sohalaridagi muammolarni hal qilishda muhim vosita bo'lib xizmat qilishi aniq.

Foydalanilgan adabiyotlar:

1. Arutyunova N. D. Tilshunoslik semantikasi: tilning sinonimik vositalari. – Moskva: Nauka, 2019.
2. Voronin S. V. Kompyuter leksikografiyasi va elektron tezauruslar yaratish muammolari. – Sankt-Peterburg: SpbDU nashriyoti, 2020.
3. Kuznetsov A. A. Tabiiy tilni avtomatik qayta ishlash: zamonaviy yondashuvlar. – Moskva: Fizmatlit, 2021.
4. Trifonov Yu. V. Mashinali o'rganish usullari va ularning ontologiyalar hamda tezauruslar yaratishdagi o'rni. – Qozon: Qozon universiteti, 2022.
5. Fellbaum C. WordNet: Elektron leksik ma'lumotlar bazasi. – Kembrij: MIT Press, 1998.
6. Jurafsky D., Martin J. Nutq va tilni qayta ishlash. – 3-nashr, Prentice Hall, 2022.
7. Yusupov A. O'zbek tilida kompyuter lingvistikasi va avtomatlashtirilgan tezauruslar yaratish tamoyillari. – Toshkent: Fan, 2023.

DIGITAL MEDIA

TIL O'RGANISHNING KOMMUNIKATIV KOMPETENSIYASI SIFATIDA

Jahongirova Iqbol Jalil qizi,

O'zMU Jurnalistika kafedrası 1-bosqich magistranti

Annotatsiya: Zamonaviy raqamli texnologiyalar ta'lim jarayoniga kiritilgan o'zgarishlar bilan birga, chet tillarini o'rganishning samaradorligini oshirishda muhim rol o'ynamoqda. Ushbu maqolada raqamli media vositalarining (forumlar, bloglar, ijtimoiy tarmoqlar, sun'iy intellekt texnologiyalari) til o'rganishda kommunikativ kompetensiyani shakllantirishdagi o'rni va ularning amaliy yondashuvlari tahlil qilingan. Raqamli texnologiyalar orqali o'rganuvchilarga tilni real kommunikativ vaziyatlarda qo'llash imkoniyatlari yaratiladi, bu esa o'quvchilarni mustaqil fikrlash va o'z til ko'nikmalarini rivojlantirishga rag'batlantiradi.

Kalit so'zlar: raqamli media, raqamli texnologiya, sun'iy intellekt, kommunikativ kompetensiya, ijtimoiy tarmoq, blog.

Аннотация: Современные цифровые технологии играют важную роль в улучшении эффективности изучения иностранных языков, благодаря изменениям, внесенным в образовательный процесс. В данной статье анализируется роль цифровых медиа-ресурсов (форумы, блоги, социальные сети, технологии искусственного интеллекта) в формировании коммуникативной компетенции при изучении языков, а также их практические подходы. Через цифровые технологии учащимся предоставляется возможность использовать язык в реальных коммуникативных ситуациях, что стимулирует их к самостоятельному мышлению и развитию языковых навыков.

Ключевые слова: цифровые медиа, цифровые технологии, искусственный интеллект, коммуникативная компетенция, социальная сеть, блог.

Abstract: Modern digital technologies play a significant role in enhancing the effectiveness of language learning through changes introduced into the educational process. This article analyzes the role of digital media tools (forums, blogs, social networks, artificial intelligence technologies) in developing communicative competence in language learning, as well as their practical approaches. Digital technologies provide learners with the opportunity to apply the language in real communicative situations, which encourages independent thinking and the development of language skills.

Keywords: digital media, digital technology, artificial intelligence, communicative competence, social network, blog.

Ta'limda raqamli texnologiyalar, ayniqsa chet tili o'rganish jarayonida, yangi imkoniyatlar yaratmoqda. Ushbu maqola, raqamli media vositalari yordamida kommunikativ kompetensiyani shakllantirish va til o'rganish samaradorligini oshirishning imkoniyatlarini ko'rib chiqadi. Raqamli texnologiyalar, jumladan, onlayn platformalar, ijtimoiy tarmoqlar, mobil ilovalar va sun'iy intellekt texnologiyalari til o'rganish jarayonini yanada interaktiv va amaliy ko'rinishga olib kelmoqda. Bu jarayon o'quvchilarga faqat grammatik qoidalar va lug'at boyligini oshirish bilan cheklanmay, balki tilni real kommunikativ vaziyatlarda qo'llash ko'nikmalarini rivojlantiradi.

Mavzuga oid adabiyotlarning tahlili. Raqamli texnologiyalar va media vositalarining ta'limda qo'llanilishi bo'yicha bir qator tadqiqotlar mavjud. Misol uchun Anderson raqamli media platformalarining til o'rganishda o'quvchilarga taqdim etayotgan interaktiv tajribalarini tahlil qiladi. Uning fikriga ko'ra, bu platformalar o'quvchilarga nafaqat grammatik bilimlarni, balki amaliy muloqot ko'nikmalarini ham rivojlantirishga yordam beradi. Shuningdek, Salome sun'iy intellekt yordamida til o'rganishning yangi imkoniyatlarini o'rganadi, bu esa til o'rganuvchilarga

interaktiv va shaxsiylashtirilgan o'quv tajribasini yaratadi. Bugungi kunda forumlar, bloglar va ijtimoiy tarmoqlarda muloqot qilish o'rganuvchilarning kommunikativ kompetensiyasini shakllantirishda muhim vosita sifatida ko'riladi.

Tadqiqot metodologiyasi. Mazkur tadqiqotda asosan sifatli tadqiqot usullari qo'llanilgan. Tahlil qilish uchun o'rganish jarayoniga oid forumlar, bloglar va ijtimoiy tarmoqlarni o'rganish asosiy metod sifatida tanlangan. Bundan tashqari, til o'rganuvchilari tomonidan foydalanuvchilarga taqdim etilgan interaktiv platformalardagi faoliyatlar o'rganildi. Tadqiqot jarayonida o'rganuvchilarning tajribalari va muvaffaqiyatlari asosida tahliliy metodlar qo'llanildi.

Tahlil va natijalar. Tadqiqot natijalari raqamli platformalar orqali til o'rganuvchilarning kommunikativ kompetensiyasi rivojlanishida muhim ijobiy o'zgarishlarni ko'rsatdi. Forumlar va ijtimoiy tarmoqlar orqali o'rganuvchilar o'z fikrlarini aniq ifodalash, bahs-munozaralarda ishtirok etish va turli kommunikativ strategiyalarni qo'llash ko'nikmalarini rivojlantirdilar. Sun'iy intellekt va virtual haqiqat texnologiyalari esa til o'rganuvchilarga real hayotdagi suhbatlarni simulyatsiya qilish orqali o'z muloqot qobiliyatlarini sinovdan o'tkazishga imkon yaratdi. Shuningdek, ijtimoiy tarmoqlarda amalga oshirilgan o'zaro muloqot va savol-javoblar o'rganuvchilarning til amaliyotini yanada samarali va qiziqarli qilishga yordam berdi.

Zamonaviy dunyoda raqamli texnologiyalar hayotning deyarli barcha sohalarida tub o'zgarishlarni yuzaga keltirmoqda. Xususan, ta'lim jarayonida, ayniqsa chet tillarini o'rganish sohasida raqamli media vositalarining ahamiyati sezilarli darajada oshib bormoqda. An'anaviy darsliklar va sinf mashg'ulotlari imkoniyatlari bilan cheklanib qolmay, zamonaviy o'quvchilar va mustaqil o'rganuvchilar til o'zlashtirishda onlayn platformalar, ijtimoiy tarmoqlar, mobil ilovalar va sun'iy intellekt texnologiyalaridan faol foydalanmoqda. Bu jarayon nafaqat o'rganish samaradorligini oshiradi, balki o'quvchilarga tilni real kommunikativ vaziyatlarda qo'llash tajribasini ham beradi. Shunday ekan, ushbu maqolada raqamli media vositalari yordamida kommunikativ kompetensiyani — ya'ni tilni amaliyotda samarali va mazmunli tarzda qo'llash qobiliyatini shakllantirish imkoniyatlari, amaliy usullari va ularning o'quv jarayonidagi o'rni tahlil qilinadi.

Digital media va kommunikativ kompetensiya. Kommunikativ kompetensiya faqatgina grammatik qoidalarni bilish yoki lug'at boyligini oshirish bilan cheklanmaydi. Bu kompetensiya, avvalo, real hayotiy vaziyatlarda to'g'ri savol berish, fikr bildirish, suhbatni davom ettirish, muloqotni samarali boshqarish va suhbatdosh bilan o'zaro anglashuvga erishish qobiliyatini o'z ichiga oladi. Shunday ko'nikmalarni shakllantirish esa faqat nazariy bilimlarni o'zlashtirish orqali emas, balki amaliy muloqot tajribasi orqali ro'yobga chiqadi.

Zamonaviy raqamli media vositalari — forumlar, bloglar, ijtimoiy tarmoqlar va onlayn suhbat platformalari til o'rganish uchun tabiiy til muhitiga yaqin sharoit yaratib beradi. O'rganuvchilar ushbu platformalarda ishtirok etish orqali jonli muloqotda qatnashib, o'z til bilimlarini amaliyotda sinab ko'rish va mustahkamlash imkoniyatiga ega bo'ladilar. Forumlardagi munozaralar, bloglardagi izohlar yoki ijtimoiy tarmoqlardagi muloqotlar orqali ular nafaqat o'z fikrlarini aniq va ravon ifodalashni, balki turli kommunikativ strategiyalarni ham o'zlashtirib boradilar. Natijada, raqamli media vositalari til o'rganuvchilarning kommunikativ kompetensiyasini shakllantirishda va rivojlantirishda muhim omil sifatida maydonga chiqmoqda.

Forumlar va diskussiya platformalari. Zamonaviy raqamli makonda til o'rganuvchilar uchun forumlar va diskussiya platformalari alohida o'rin tutadi. Reddit, Duolingo Discussions kabi global forumlar, shuningdek, turli tillarga ixtisoslashgan maxsus onlayn maydonlar o'quvchilarga real mavzularda muloqot qilish, savol-javob shaklida bilim almashish va tajriba orttirish imkonini yaratmoqda. Ushbu platformalarda foydalanuvchilar kundalik hayot, madaniyat, adabiyot, fan, texnologiya kabi keng qamrovli mavzular atrofida o'z fikrlarini bildirishlari va o'zaro interaktiv suhbat olib borishlari mumkin. Forumlardagi muloqotlar o'rganilayotgan tilni faol qo'llash imkoniyatini taqdim etadi. O'quvchilar o'z fikrlarini aniq, tushunarli va mantiqiy ifodalashni, o'zgalarning fikrlariga izchil va madaniyatli tarzda javob berishni mashq qiladilar. Bundan tashqari, bahs-munozaralarda ishtirok etish orqali argumentatsiya qilish, qarama-qarshi fikrlarni

muhokama qilish va o'z pozitsiyasini himoya qilish ko'nikmalarini ham rivojlantiradilar. Bularning barchasi esa o'rganuvchilarning kommunikativ kompetensiyasini chuqur va kompleks tarzda shakllantirishga xizmat qiladi.

Ijtimoiy tarmoqlar va hamkorlik muloqoti. Bugungi raqamli asrda ijtimoiy tarmoqlar til o'rganuvchilar uchun nafaqat muloqot vositasi, balki mustaqil va hamkorlikda o'rganish imkoniyatlarini taqdim etuvchi interaktiv platformalarga aylangan. Facebook, Instagram, Twitter, Telegram kabi mashhur ijtimoiy tarmoqlar o'quvchilarga cheklovlarisiz muloqot qilish, o'z bilim va tajribalarini boshqalar bilan bo'lishish hamda doimiy ravishda til amaliyoti olib borish imkonini bermoqda. Ayniqsa, til o'rganishga ixtisoslashgan guruhlar va mavzuli chatlar til o'rganuvchilarga o'zaro fikr almashish, savol-javoblar o'tkazish va real muloqot muhitini yaratish imkoniyatini beradi. Maxsus guruhlarda har kuni yangi mavzular atrofida suhbatlar tashkil etilishi o'quvchilarning doimiy ravishda yangi so'zlar va iboralarni o'zlashtirishiga, grammatik bilimlarini amaliyotda mustahkamlashiga yordam beradi. Shuningdek, jonli efirlar (live streaming) orqali o'tkaziladigan ma'ruzalar, seminarlar va suhbatlar til o'rganuvchilarga tabiiy muloqot sharoitida ishtirok etish va eshitish-ko'rish asosida til sezgirligini oshirish imkoniyatini yaratadi.

Bundan tashqari, ijtimoiy tarmoqlarda o'zbek tilida, ingliz tilida yoki boshqa tillarda o'zlari o'rganayotgan manbalarni ulashishlari, ularni birgalikda tahlil qilib solishtirishlari yoki ularga izohlar berish orqali ham yozma nutq ko'nikmalari rivojlanadi. Muloqot jarayonida foydalanuvchilar grammatika va uslubiy xatolarini amaliy tarzda ko'rib chiqadilar va o'z ustida ishlash imkoniyatiga ega bo'ladilar. Shu orqali ular nafaqat til bilimlarini kengaytiradilar, balki madaniyatlararo kommunikatsiya qobiliyatlarini ham shakllantiradilar. Umuman olganda, ijtimoiy tarmoqlarda olib boriladigan hamkorlik muloqoti til o'rganuvchilar uchun muntazam, jonli va sarmazmun til amaliyotini ta'minlaydi, bu esa ularning kommunikativ kompetensiyasini uzluksiz rivojlantirib borishga xizmat qiladi.

Kompyuter texnologiyalari va sun'iy intellekt. Zamonaviy raqamli texnologiyalar, ayniqsa kompyuter tizimlari va sun'iy intellekt (AI) asosidagi vositalar til o'rganish jarayonini tubdan o'zgartirib, uni yanada samarali va shaxsiylashtirilgan ko'rinishga olib kelmoqda. Duolingo, Babel, Memrise kabi mashhur ilovalar foydalanuvchilarga individual ehtiyojlariga moslashtirilgan o'quv rejalarini taqdim etadi. Ushbu dasturlar foydalanuvchilarning bilim darajasini aniqlab, xatoliklarini avtomatik tarzda tahlil qiladi va ularga yaxshilanish uchun aniq tavsiyalar beradi. Natijada, har bir o'quvchi o'zining zaif tomonlariga ko'proq e'tibor qaratib, til o'rganish jarayonida individual yondashuv asosida rivojlana oladi. Bundan tashqari, sun'iy intellekt yordamida ishlab chiqilgan chatbotlar va virtual suhbatdoshlar foydalanuvchilarga real hayotga yaqin sharoitda suhbat olib borish imkoniyatini yaratmoqda. Masalan, til o'rganuvchi AI chatbot bilan turli mavzularda savol-javob qilishi, suhbat davom ettirishi va o'z muloqot qobiliyatini real vaqtda sinovdan o'tkazishi mumkin. Bu esa, ayniqsa, chet tilida gapirishga qiyinchilik sezadigan o'rganuvchilar uchun o'z ishonchini oshirish va nutqdagi tabiiylikni rivojlantirishda juda muhim ahamiyat kasb etadi. Shuningdek, virtual haqiqat (VR) va kengaytirilgan haqiqat (AR) texnologiyalari til o'rganish jarayonini yanada jonlantirib, uni immersiv (to'liq sho'ng'ish) tajribaga aylantirmoqda. Masalan, VR texnologiyalari yordamida o'quvchi o'zini xorijiy mamlakat ko'chalarida yoki universitet auditoriyasida his qilishi, mahalliy aholi bilan ingliz, nemis, ispan yoki boshqa tillarda muloqot qilayotgandek tajriba orttirishi mumkin. AR texnologiyalari esa o'quvchining atrof-muhitiga virtual elementlarni qo'shib, real hayotdagi ob'ektlar bilan chet tilida interaktiv muloqot qilish imkonini beradi.

Bu kabi innovatsion texnologiyalar til o'rganishni nafaqat samaraliroq, balki ko'proq motivatsion va qiziqarli jarayonga aylantiradi. Ayniqsa, amaliy nutq ko'nikmalarini rivojlantirish va kommunikativ kompetensiyani shakllantirishda bunday texnologik vositalarning o'rni beqiyosdir. Shaxsiy tajribaga asoslangan mashg'ulotlar orqali o'rganuvchilar nafaqat til bilimlarini kengaytiradilar, balki global raqamli jamiyatda muvaffaqiyatli muloqot qilish uchun zarur bo'lgan ko'nikmalarni ham rivojlantiradilar.

Wattpad, Tumblr va Inkitt kabi platformalar. Bugungi raqamli transformatsiya davrida an'anaviy ta'lim usullari zamonaviy texnologiyalar bilan chambarchas bog'liq holda rivojlanmoqda. Til o'rganuvchilar endilikda nafaqat darslik va lug'atlarga tayana boshlashdi, balki turli xil onlayn platformalar orqali amaliy ko'nikmalarini mustahkamlash, til muhitiga kirib borish va kommunikativ kompetensiyalarini rivojlantirish imkoniga ega bo'ldilar. Ayniqsa, Wattpad, Tumblr va Inkitt kabi yozuv va ijodga ixtisoslashgan raqamli maydonlar til o'rganish jarayonida samarali yordamchi vosita sifatida ajralib turmoqda. Wattpad — bu butun dunyo bo'ylab millionlab foydalanuvchilarni o'z atrofida to'plagan va ular orasida ijod mahsulotlarini baham ko'rishga imkoniyat yaratgan platformadir. Til o'rganuvchilar uchun Wattpad kichik hikoyalar, esse va romanlar yozish orqali o'z yozma ko'nikmalarini sinovdan o'tkazish va rivojlantirish uchun qulay maydon hisoblanadi. Platformadagi faol auditoriya foydalanuvchilarning asarlarini sharhlab, xatolarni ko'rsatib, stilistika va grammatika borasida amaliy tavsiyalar beradi. Shu tarzda, o'rganayotgan shaxslar nafaqat nazariy bilimlarini mustahkamlashadi, balki muloqotdagi nozik ifoda uslublarini ham puxta o'zlashtiradilar. Wattpad'dagi janr va til asosida shakllangan ixtisoslashgan guruhlar esa o'z navbatida o'rganuvchining tanlagan til sohasida chuqurroq ishlashga imkon yaratadi.

Tumblr platformasi esa blog shaklida qisqa postlar, uzun maqolalar, hamda vizual va matnli kontentlar almashinuvi uchun mo'ljallangan bo'lib, foydalanuvchilarni qisqa, aniq va ta'sirchan fikr bildirish ko'nikmasini rivojlantirishga undaydi. Til o'rganish jarayonida Tumblr bloggerlarining postlarini o'qish, sharhlash va ular bilan muloqot qilish orqali foydalanuvchilar jonli va zamonaviy tildan foydalanish imkoniyatiga ega bo'ladilar. Platformadagi hashtaglar va mavzular asosidagi guruhlar orqali esa ma'lum bir soha yoki qiziqish doirasida faol muloqot olib borish, shu orqali so'z boyligi va ifoda erkinligini oshirish mumkin. Ayniqsa, ingliz, fransuz va ispan tillarida faol bo'lgan jamoalar til o'rganuvchilar uchun amaliy mashg'ulot vazifasini o'tashi mumkin.

Inkitt esa asosan adabiy asar yozishga qiziquvchi ijodkorlar uchun mo'ljallangan platforma bo'lib, foydalanuvchilarga o'z hikoya va romanlarini keng auditoriyaga taqdim etish imkonini beradi. Til o'rganishda Inkitt alohida ahamiyat kasb etadi, chunki bu platforma grammatik jihatdan to'g'ri yozish bilan birga stilistik va badiiy ifoda uslublarini ham chuqur o'zlashtirishga undaydi. Foydalanuvchilar o'z asarlaridagi xatolarni ko'rib chiqish, tahrir qilish va takomillashtirish orqali yozma ko'nikmalarini bosqichma-bosqich rivojlantiradilar. Shuningdek, ular boshqa foydalanuvchilarning asarlarini o'qish orqali ham o'z ifoda imkoniyatlarini kengaytiradilar.

Mazkur platformalarning umumiy jihati shundaki, ular foydalanuvchini passiv o'rganishdan faol muloqot va amaliy mashq qilish sari yo'naltiradi. Til o'rganish jarayonida xatolardan qo'rqqanlik, tajriba orttirish va mustahkam til kompetensiyasini shakllantirishda aynan bunday ochiq va jonli platformalar katta yordam beradi. Kichik hikoya, post yoki maqola yozish orqali o'zini erkin ifoda etishni o'rganayotgan o'rganuvchilar asta-sekin mukammal va o'ziga ishongan yozuvchilarga aylanishadi. Bundan tashqari, Wattpad, Tumblr va Inkitt kabi global platformalarda turli mamlakatlardan bo'lgan foydalanuvchilarning ishtiroki mavjudligi til o'rganuvchilarga turli madaniy kontekstlarni anglab yetish, turli xalqlarga xos ifoda shakllarini o'rganish va madaniyatlararo muloqot ko'nikmasini rivojlantirish imkonini beradi. Bu esa o'z navbatida nafaqat til bilimini, balki global tafakkur va dunyoqarashni ham shakllantiradi.

Bloglar va yozma ifoda ko'nikmalari. Til o'rganishda blog yuritish samarali va ijodiy yondashuvlardan biridir. O'rganuvchilar shaxsiy bloglarida kundalik hayotiy tajribalari, kitob sharhlari yoki o'zlari qiziqqan mavzular haqida maqolalar yozish orqali yangi so'z va iboralarni faol o'zlashtiradilar. Har bir post yozish jarayonida ular matn tuzilishiga, grammatik to'g'rilikka va ifoda uslubiga e'tibor qaratishga majbur bo'ladilar, bu esa yozma nutqni sezilarli darajada rivojlantiradi.

Bundan tashqari, til o'rganuvchi nafaqat o'z blogini yuritishi, balki boshqa bloglarni kuzatishi ham juda foydalidir. Boshqalar tomonidan yozilgan postlarni o'qish, ularga izohlar yozish va muloqotga kirishish orqali o'quvchi o'zining o'qish, tahlil qilish va tanqidiy fikrlash

qobiliyatlarini rivojlantiradi. Ayniqsa, har xil bloglar turli auditoriyalarni bir joyga jamlagani uchun o'rganish imkoniyatlari yanada kengayadi.

Masalan, **Reddit** foydalanuvchilari ko'proq muhokama va savol-javoblarga e'tibor qaratadilar, **Wattpad** foydalanuvchilari esa hikoya va badiiy matnlar orqali ifoda uslubiga urg'u beradilar. **Tumblr** esa qisqa blog yozuvlari va vizual kontent bilan tilni qisqa shakllarda ifodalashni rivojlantiradi. Blog maydonlari shu kabi tor yo'nalishdagi auditoriyalarni birlashtirib, umumiy kommunikativ platformaga aylanishi mumkin. Natijada til o'rganuvchi nafaqat biror aniq uslub bilan cheklanmay, balki turli muloqot uslublarini, janr va uslubiy xususiyatlarni ham o'zlashtirib boradi. Bu kabi integratsiyalashgan blog muhitida qatnashish tilni faqat nazariy bilish bilan cheklanib qolmay, uni amaliyotda qo'llash, kontekstga mos ifoda qilish va muloqotni boshqarish kabi chuqurroq ko'nikmalarni ham shakllantiradi. Natijada o'rganuvchilar yozma va og'zaki ifoda jihatidan mustahkam kommunikativ kompetensiyaga ega bo'lib boradilar.

Xulosa. Zamonaviy raqamli media vositalarining til o'rganish jarayonidagi roli kundan-kunga ortib bormoqda. Wattpad, Tumblr va Inkitt kabi platformalar o'rganuvchilarga nafaqat grammatika va lug'at boyligini oshirish, balki yozma va og'zaki muloqot kompetensiyalarini ham rivojlantirish imkonini bermoqda. Ushbu platformalar tilni tabiiy muhitda qo'llash, real muloqot sharoitida fikr almashish va madaniy kontekstda o'rganish imkoniyatini yaratib, o'rganish jarayonini interaktiv, amaliy va qiziqarli shaklga olib chiqmoqda. Natijada, o'rganuvchilar nafaqat til bilimini mustahkamlash, balki global tafakkur va madaniy ongni rivojlantirish imkoniga ega bo'lmoqdalar.

Takliflar. Til o'rganish jarayonida raqamli platformalardan foydalanishni yanada kengaytirish va maqsadli yo'naltirish zarur. Ta'lim muassasalarida Wattpad, Tumblr, Inkitt kabi ijodiy platformalardan amaliy mashg'ulotlarda foydalanish, o'rganuvchilarni o'z fikrlarini erkin ifodalashga va tahliliy tafakkurini rivojlantirishga undaydi.

Bundan tashqari, sun'iy intellekt va virtual haqiqat texnologiyalarini til o'rganish jarayoniga integratsiya qilish orqali, o'quvchilarni real muloqotga yaqin muhitda mashq qilish imkoniyatlari bilan ta'minlash tavsiya etiladi.

Shuningdek, bloglar va ijtimoiy tarmoqlar orqali o'rganuvchilarning mustaqil fikrlash, yozma va og'zaki ifoda ko'nikmalarini rivojlantirish uchun sharoit yaratish zarur. Fikr almashish, izohlar yozish va kontent yaratish amaliyoti til o'zlashtirish jarayonini yanada jonlantiradi.

Til o'rganishda madaniy almashinuv va global tafakkurni shakllantirishga alohida e'tibor qaratish kerak. Chunki raqamli media vositalari orqali turli xalqlarning madaniyatlari bilan tanishish, xalqaro muloqot madaniyatini rivojlantirish va til o'rganishni yanada chuqurroq va mazmunli qilish imkoniyati yaratiladi.

Adabiyotlar ro'yxati:

1. Anderson, J. (2016). *Digital Media and Language Learning: Theory, Research and Practice*. Routledge.
2. Harrison, D. (2020). *Interactive Communication through Social Media in Language Education*. Journal of Educational Technology.
3. Salomé, A. (2018). *AI and Language Learning: Enhancing Interactive and Personalized Learning*. Educational Innovation Journal.

SUN'IY INTELLEKT YORDAMIDA TIL O'RGANISH: TEXNOLOGIYALAR TA'LIMYIY SAMARADORLIGI VA INSON OMILI MUVOZANATI

Omonov Ravshanjon

O'zMU Jurnalistika kafedrası dotsenti,
siyosiy fanlar bo'yicha falsafa doktori

Annotatsiya: Sun'iy intellekt texnologiyalarining til o'rganishdagi imkoniyatlari va chegaralari tahlil qilingan. Globallashuv va raqamli transformatsiya jarayonida til o'rganish shaxs rivoji va mehnat bozorida raqobatbardoshlikni oshirish uchun muhim omilga aylanmoqda. Yaqin yillarda SI platformalari, masalan, Duolingo, ChatGPT va Replika kabi vositalar, an'anaviy ta'lim metodlarini butunlay yangicha shaklda taklif etmoqda. Bu o'rganish jarayonini tezlashtirish, motivatsiyani oshirish va individuallashtirilgan o'rganish imkoniyatlarini yaratishda muhim ahamiyatga ega. Ammo, bunday platformalardan foydalanishdagi asosiy muammo shundaki, ular ijtimoiy va madaniy bilimlarni rivojlantirish, hayotiy mo'jarolar va empatiyani shakllantirishda kamchiliklarga olib keladi. Shu sababli, ushbu tadqiqotda SI platformalari orqali mustaqil o'rgangan va an'anaviy ta'lim olgan guruhlarining bilimi, fikrlash qobiliyati va motivatsiyasi o'rganilib, samaradorlik ko'rsatkichlari to'g'risidagi ma'lumotlar taqqoslangan. Natijalar shuni ko'rsatdiki, SI platformalari grammatik va leksik bilishda muvaffaqiyatli bo'lsa-da, real muloqot va ijtimoiy adaptatsiyada muammolar mavjud. Bu esa, inson omilining, xususan, o'qituvchining roli ta'limdagi samaradorlik uchun ahamiyatli ekanini ko'rsatadi.

Kalit so'zlar: sun'iy intellekt, til o'rganish, an'anaviy ta'lim, motivatsiya, grammatika, ijtimoiy kompetensiyalar, kommunikatsiya.

Аннотация: В данной статье анализируются возможности и ограничения технологий искусственного интеллекта (ИИ) в процессе изучения языков. В условиях глобализации и цифровой трансформации изучение языков становится важным фактором личностного развития и повышения конкурентоспособности на рынке труда. В последние годы такие ИИ-платформы, как Duolingo, ChatGPT и Replika, предлагают альтернативные подходы к традиционным методам обучения. Эти технологии играют важную роль в ускорении процесса обучения, повышении мотивации и создании индивидуализированных условий для изучения. Однако использование таких инструментов имеет ограничения в развитии социокультурных компетенций, навыков ориентации в жизненных ситуациях и эмпатии. В исследовании сравнивались знания, мышление и мотивация групп, обучавшихся с помощью ИИ и традиционными методами. Результаты анализа показали, что ИИ-платформы эффективны в развитии грамматических и лексических знаний, однако имеют недостатки в реальной коммуникации и социальной адаптации. Это подтверждает значимость человеческого фактора, в частности, роли преподавателя, в обеспечении эффективности образования.

Ключевые слова: искусственный интеллект, изучение языков, традиционное обучение, мотивация, грамматика, социальные компетенции, коммуникация.

Abstract: This article analyzes the opportunities and limitations of artificial intelligence (AI) technologies in language learning. In the context of globalization and digital transformation, learning languages is becoming a key factor in personal development and increasing competitiveness in the labor market. In recent years, AI platforms such as Duolingo, ChatGPT, and Replika have been offering alternative approaches to traditional educational methods. These technologies play an important role in accelerating the learning process, enhancing motivation, and enabling personalized learning experiences. However, their use reveals limitations in developing socio-cultural competence, real-life decision-making, and empathy. This study

compares the knowledge, thinking abilities, and motivation of groups who learned independently through AI tools versus those who received traditional education. The results show that while AI platforms are effective in developing grammar and vocabulary skills, they struggle with real communication and social adaptation. This highlights the essential role of human factors, especially the teacher's involvement, in achieving educational effectiveness.

Keywords: artificial intelligence, language learning, traditional education, motivation, grammar, social competences, communication.

Globallashuv jarayoni va raqamli texnologiyalarning jadal sur'atlarda rivojlanishi natijasida chet tilini o'zlashtirish shaxsning intellektual salohiyatini yuksaltirish, madaniy gorizontlarini kengaytirish hamda mehnat bozorida raqobatbardoshligini ta'minlashning asosiy omillaridan biriga aylanmoqda. Ayniqsa, so'nggi yillarda sun'iy intellekt (SI) texnologiyalari asosida ishlab chiqilgan ta'limiy platformalar — **Duolingo, ChatGPT, Replika** va shunga o'xshash zamonaviy dasturlar — til o'rganish jarayoniga mutlaqo yangi, innovatsion yondashuvlarni taklif qilmoqda. Mazkur platformalar an'anaviy o'qitish usullariga nisbatan bir qator afzalliklarga ega: ular orqali o'quvchiga individual yondashuv asosida tezkor, aniq va avtomatlashtirilgan fidbek taqdim etiladi, foydalanuvchi ehtiyojlariga mos ravishda moslashadigan o'quv modullari va mashqlar taklif etiladi. Shuningdek, ushbu vositalarda mavjud bo'lgan interaktiv muhit, vizual va audio vositalarning keng qo'llanilishi tilni o'rganish jarayonini nafaqat samarali, balki qiziqarli va motivatsion jihatdan rag'batlantiruvchi holga keltiradi. Qolaversa, bu kabi platformalarning to'liq onlayn shaklda ishlashi sababli, istalgan vaqt va joyda ta'lim olish imkoniyati vujudga keladi, bu esa mustaqil o'quv jarayonini qo'llab-quvvatlaydi. Ayniqsa, zamonaviy yoshlar orasida bu shakldagi ta'limga qiziqishning ortib borishi ham bugungi raqamli davrda SI texnologiyalarining naqadar dolzarbligini ko'rsatadi. Biroq, ushbu texnologiyalardan muntazam, ya'ni nazoratsiz va to'g'ri pedagogik yo'naltirishsiz foydalanish ayrim jiddiy muammolarni keltirib chiqarishi mumkin. Jumladan, tilni chuqur o'rganishga zarur bo'lgan **mustaqil fikrlash, tanqidiy tahlil qobiliyati, real muloqot orqali ijtimoiy tajriba to'plash**, hamda **jamiyatga moslashish (ijtimoiy adaptatsiya)** kabi muhim kompetensiyalar yetarli darajada shakllanmasligi xavfi mavjud. Chunki sun'iy intellekt vositalari ko'p hollarda ma'lumotni avtomatik tarzda yetkazish va testlarga asoslangan fidbek berish bilan cheklanadi. Inson omili, ya'ni jonli muloqot, ehsosiy munosabat, ijtimoiy signallarni qabul qilish kabi jarayonlar esa chetda qolib ketadi. Shu sababli, ushbu maqolada sun'iy intellektga asoslangan platformalar orqali mustaqil ravishda til o'rganayotgan o'quvchilar bilan an'anaviy dars jarayonida o'qituvchi rahnamoligida ta'lim olayotgan o'quvchilar o'rtasidagi farqlar atroflicha tahlil qilinadi. Xususan, **bilim chuqurligi, fikrlash qobiliyatining rivojlanish darajasi**, hamda **o'qishga bo'lgan ichki motivatsiyadagi farqlar** o'rganiladi. Bu tahlil natijalari orqali sun'iy intellekt vositalarining imkoniyatlari va cheklovlari aniqroq namoyon bo'ladi. Shu bilan birga inson omilining ta'lim jarayonidagi o'rnini baholash imkoni paydo bo'ladi.

Adabiyotlar tahlili. Sun'iy intellektning ta'limdagi roli bo'yicha olib borilgan tadqiqotlar, asosan, SI texnologiyalarining til o'rganish jarayonidagi samaradorligini ko'rsatmoqda. M.Azimova va X.Reynders o'z asarlarida Duolingo kabi platformalarning grammatika va lug'atni mustahkamlashdagi foydaliligini ta'kidlaganlar. Biroq, R.Ustyujanina va Liu bu texnologiyalarni tanqid qilib, inson omillari – empatiya, madaniy kontekst va jonli muloqotning yetishmasligini ta'kidlagan. Bularni hisobga olgan holda, ushbu ishda SI vositalarining ta'limdagi muvaffaqiyati va ularning cheklovlari tahlil qilinadi.

Tadqiqot metodologiyasi. Tadqiqotda ikkita guruh ishtirok etdi: biri SI platformalaridan foydalangan holda mustaqil ta'lim olganlar, ikkinchisi esa an'anaviy darslarda o'qituvchi ishtirokida o'rganganlar. Har bir guruhda 15 nafardan o'quvchi bo'lib, ular ikki oy davomida bir xil mavzular asosida baholandi. Tahlilda grammatika, muloqot va motivatsiya darajalari tekshirildi.

Tahlil va natijalar. Tadqiqot natijalari shuni ko'rsatadiki, SI guruhida grammatika va lug'at bo'yicha yuqori natijalar qayd etildi, lekin real muloqot va ijtimoiy adaptatsiya bo'yicha qiyinchiliklar kuzatildi. An'anaviy guruhda esa til orqali fikr ifodasi va muloqotda yuqori ko'rsatkichlar qayd etilgan, ammo grammatik xatolar ko'proq bo'lgan. Motivatsiya darajasi SI guruhida dastlab yuqori bo'lsada, vaqt o'tishi bilan pasaydi, an'anaviy guruhda esa motivatsiya barqaror saqlanib qoldi.

Xulosa va takliflar. Tadqiqot natijalari shuni ko'rsatadiki, sun'iy intellekt vositalari til o'rganishda samarali bo'lishi mumkin, ayniqsa mustaqil ta'lim shaklida. Biroq, ular ijtimoiy va kommunikativ kompetensiyalarni shakllantirishda yetarli emas. Shu tufayli, eng samarali yondashuv – gibril model bo'lib, an'anaviy ta'limni sun'iy intellekt asosidagi vositalar bilan birgalikda qo'llashdir. O'qituvchilar SI vositalarini yordamchi resurs sifatida foydalanishi, ammo jonli muloqotni saqlab qolgan holda o'quvchini faol ishtirokchi sifatida shakllantirishga harakat qilishi lozim. Kelajakda ta'lim texnologiya va inson o'rtasidagi hamkorlikda rivojlanishi zarur.

Sun'iy intellekt vositalarining til o'rganishdagi imkoniyatlari. Bugungi raqamli asrda ta'lim sohasi jadal sur'atlar bilan o'zgarib, sun'iy intellekt (SI) texnologiyalari bu jarayonning markaziy elementiga aylanmoqda. Ayniqsa, chet tilini o'zlashtirish jarayonida SI vositalari innovatsion yondashuvlarni taklif qilib, an'anaviy usullardan farq qiluvchi, yangi, samarali modellarni yaratishga xizmat qilmoqda. Bu kabi texnologiyalar faqat ma'lumot berish bilan cheklanib qolmasdan, shu ma'lumotni qayta ishlash, tahlil qilish, foydalanuvchiga moslashish va o'quv jarayonini shaxsiylashtirish imkonini ham taqdim etadi.

Xususan, **Natural Language Processing (NLP)** — tabiiy tilni qayta ishlash texnologiyalari — orqali ishlovchi platformalar o'quvchining lug'aviy zahirasi, grammatik bilimi va tilda ifoda qilish qobiliyatini avtomatik tarzda tahlil qilib, ular asosida individuallashtirilgan o'quv kurslarini taklif etadi. Bu esa an'anaviy "barchaga bir xil" ta'lim modelidan farqli ravishda, har bir foydalanuvchining bilimi, o'rganish sur'ati va ehtiyojlariga moslashadigan dinamik tizimni yuzaga keltiradi.

Duolingo platformasi, masalan, interaktiv mashqlar to'plamini avtomatik ravishda foydalanuvchining bilim darajasiga mos ravishda tanlaydi. U o'quvchining xatolarini tahlil qiladi, eng ko'p qiyinchilik tug'dirgan nuqtalarda ko'proq mashqlar taklif qiladi va muvaffaqiyatli bosqichlarni doimiy ravishda rag'batlantiradi. Bu jarayon nafaqat bilimni mustahkamlash, balki o'rganishga bo'lgan qiziqishni ham kuchaytiradi.

ChatGPT kabi til modeli esa real vaqtda muloqot muhitini simulyatsiya qiladi. U foydalanuvchi bilan sun'iy suhbat olib borish orqali nutqning erkinligini, izchilligini va ma'nolilik darajasini oshirishga ko'maklashadi. Buning o'zi til o'rganishda muhim sanaladi, chunki jonli muloqot ko'nikmasi til bilishning eng yuqori bosqichlaridan biri hisoblanadi. ChatGPT foydalanuvchining savollariga javob beradi, tushunmovchiliklarni izohlaydi va hatto yangi mavzular yuzasidan mulohaza qilishga undaydi.

Replika kabi dasturlar esa sun'iy intellekt orqali inson bilan empatik muloqotga kirishadi. Ushbu platformada asosiy e'tibor faqat grammatika yoki lug'atga emas, balki muloqotdagi hissiy jihatlarini tushunish, munosabat ifoda qilish, ma'naviy qo'llab-quvvatlash kabi omillarga qaratiladi. Bu esa real hayotdagi ijtimoiy muloqot ko'nikmalarini shakllantirishga xizmat qiladi, ayniqsa yosh o'quvchilar va ijtimoiy izolyatsiyaga moyil foydalanuvchilar uchun muhimdir.

Shu tariqa, sun'iy intellekt vositalari orqali taqdim etilayotgan imkoniyatlar nafaqat o'qitish jarayonini avtomatlashtirish, balki uni individuallashtirish, qiziqarli qilish va psixologik jihatdan qulay muhit yaratishga qaratilgandir. Ular orqali o'quvchilar nafaqat mustaqil o'rganish, o'qishni istalgan vaqt va joyda amalga oshirish imkoniyatiga ham ega bo'ladi. Bundan tashqari, sun'iy intellekt platformalaridagi geymifikatsiya, ya'ni o'qishni o'yinga asoslash, o'quvchi motivatsiyasini oshirishda muhim rol o'ynaydi.

Shu bilan birga, mazkur vositalar ta'limda yangicha imkoniyatlarni ochib berishi bilan birga, inson omilining o'rnini to'liq bosolmasligini ham unutilmasligi lozim.

Bu vositalar, albatta, individual o'rganish uchun samarali va qulay ko'rinish bermoqda, chunki ular foydalanuvchiga moslashib, yagona o'quv darajasiga ega bo'lgan shaxsga mo'ljallangan kurslar va mashqlar taklif qiladi. Bunday tizimlar o'qishni tezlashtirish va shaxsning yozish, o'qish, tinglash kabi asosiy til ko'nikmalarini samarali rivojlantirishga xizmat qiladi. Ammo, ushbu texnologiyalar chuqur ta'limiy yondashuv va ijtimoiy integratsiyani ta'minlashda bir qator chegaralarga ega ekanligi seziladi.

Ta'lim jarayonida axborotni faqat avtomatik baholashga asoslangan tarzda yoritish bilan chegaralanish mumkin, lekin bu metodika jamiyatdagi aniq munosabatlarni, madaniy kontekst, psixologik holatni va jonli muloqotni inobatga olmasligi mumkin. Masalan, **empatiya** kabi muhim komponentlar, o'quvchi bilan haqiqiy hamkorlik, va yordam berish jarayonlari ko'p hollarda sun'iy intellekt tarafidan o'ziga xos tarzda qayta ishlanmaydi. Bundan tashqari, **muloqotdagi adashishlar** yoki **ijtimoiy signallar** ham sun'iy intellekt yordamida o'zlashtirilgan ta'lim jarayonida to'liq aniqlanmasa, o'quvchining haqiqiy bilimlarini tekshirish va ularni chuqurroq tushunishda muammolar tug'ilishi mumkin.

Ijtimoiy adaptatsiya va madaniy integratsiya jarayonida inson omili yordam bermasligi sababli, interaktiv va empatik muloqotlarni tashkil etish imkoni cheklanadi. Bu esa, ayniqsa, tilni o'rganayotgan shaxs uchun muhim – shu bilan birga, o'quvchilar o'zlarining yoki muhitidagi madaniy va ijtimoiy belgilarni tushunishda kamchiliklarga duch keladi. **Jonli muloqot, muloqotdagi ishonch va psixologiya**, hamda **jamiyatdagi ijtimoiy munosabatlar** ko'nikmalari ishlanmagan holda, bu turdagi texnologik vositalar o'zining eng samarali foydalanish sohasida chegaralanadi.

Sun'iy intellekt va an'anaviy ta'lim: solishtiruvchi tahlil. Mazkur tadqiqotda ikki turdagi o'quvchilar ishtirok etdi. Birinchi guruh — **sun'iy intellekt platformalari** orqali mustaqil o'rganishga asoslangan yoshlar, ikkinchi guruh esa **an'anaviy ta'lim usullari** orqali, ya'ni o'qituvchi bilan jonli darslarda ishtirok etgan o'quvchilardan iborat bo'ldi. Har ikki guruhda 15 nafardan o'quvchi ishtirok etdi va ularning har biri o'qish jarayonida bir xil mavzular va baholash mezonlari asosida tekshirildi.

Tadqiqot davomida, har ikki guruhga bir xil mavzular taqdim etilib, ularning til bilimi, muloqotdagi ishlash qobiliyatlari, grammatik qoidalarga nisbatan aniqlik va nutqdagi erkinlik darajasi o'rganildi. O'quvchilar o'rganishda istifodalangan vosita va metodlar asosida turli ishlanmalar va mashqlar bajarishgan.

Sun'iy intellekt platformalari orqali o'rgangan guruhda natijalar har bir o'quvchining shaxsiy darajasiga mos keladigan, avtomatik tarzda berilgan va tezkor baholashga asoslangan mashqlarga qaratilgan. Bu guruhda grammatik va leksik mashqlarda yuqori natijalarga erishildi, lekin real muloqot va ijtimoiy adaptatsiya jarayonlari ko'proq muammolarni keltirib chiqardi. Texnologiyalar to'liq ijtimoiy kontekstni hisobga olmaganligi uchun foydalanuvchilarning ma'naviy va emotsional yondoshuvlari yetarli darajada rivojlanmadi.

An'anaviy ta'limni olayotgan guruhda esa o'qituvchi bilan yaqin muloqot, hayotga yaqin vaziyatlar va ijtimoiy ko'nikmalarga e'tibor qaratilgan. Bu guruhning natijalari muloqotda faol bo'lish, ijtimoiy interaksionlar va mavzu bo'yicha fikr bildirish qobiliyatida yuqori bo'ldi. Shunga qaramasdan, grammatik va leksik xatolar ko'p bo'lgani sababli, ularning tildagi aniqlik darajasi sun'iy intellekt orqali o'rganganlardan pastroq bo'ldi.

Tadqiqot natijalari shundan iboratki, **SI platformalari** orqali til o'rgangan guruh grammatik topshiriqlarni bajarishda yuqori natijalar ko'rsatdi. Biroq, bu guruhda **real muloqot va ijtimoiy holatlarda** qiyinchiliklar kuzatildi. O'quvchilar asosan avtomatik tarzda taklif qilingan mashqlarni bajargan bo'lsa-da, ular orasida ijtimoiy kontekstda muloqot qilishda muammolar yuzaga keldi. Ular, asosan, grammatik jihatdan to'g'ri javoblar bera olishsa ham, real hayotdagi suhbatlarda bir-biri bilan o'zaro tushunish va o'z fikrlarini jonli ifodalashni qiyinlashtirishdi. Bu, avvalo, **sun'iy intellekt vositalarining** inson omillariga asoslangan kommunikativ komponentlarni yetarli darajada ta'minlamasligidan kelib chiqmoqda.

An'anaviy ta'lim guruhida esa, aksincha, muloqot tajribasi, til orqali fikr ifodai va **ijtimoiy faollik** yuqori bo'ldi. O'quvchilar bir-biri bilan jismoniy muloqotga kirishib, o'zaro fikr almashish va muhokama qilish orqali tilni yanada chuqurroq o'zlashtirishdi. Ular orasida **ijtimoiy faoliyati** va tilni foydalanishdagi **madaniy kontekstni tushunish**da sezilarli muvaffaqiyatlar qayd etildi. Ammo, ushbu guruhda grammatik xatolar nisbatan ko'proq bo'lishi mumkin edi, chunki o'qituvchi bilan real vaqtda muloqot davomida ayrim grammatik qoidalar va strukturaviy jihatlar aniqlanmagan yoki to'liq tushunilar va ko'rsatilmagan bo'lishi mumkin.

Motivatsiya jihatidan, **SI platformalari** guruhida dastlab motivatsiya juda yuqori edi, chunki o'quvchilar qisqa muddat ichida avtomatik tarzda natijalar ko'rishdi va tizim doimiy ravishda o'zgarib turishini va yangi mashqlar bilan qiziqtirib turdi. Ammo vaqt o'tishi bilan motivatsiya pasayib, o'quvchilar uchun ilgari surilgan maqsadlar va natijalar bir xil ko'rinishda bo'lib, yangi qiziqishlarni uyg'otishning samarali mexanizmlari yo'qligi kuzatildi.

An'anaviy guruhda esa motivatsiya barqaror saqlanib qolgan, chunki o'qituvchi tomonidan doimiy rag'batlantirish, qiyinchiliklarni yengib o'tishda yordam berish va muvaffaqiyatlar va xatolarni real vaqtda muhokama qilish jarayoni o'quvchilarni davom etishga rag'batlantirdi.

Bu holat, shubhasiz, **SI vositalarining texnik jihatdan samarali** bo'lishiga qaramay, **inson omillari** — ayniqsa, o'qituvchining roli — ta'limda hal qiluvchi ahamiyatga ega ekanligini ko'rsatadi. **O'qituvchining** emotsional qo'llab-quvvatlashi, rag'batlantirishi va tahliliy yondashuvi o'quvchilarning motivatsiyasini, ularning o'rganishga bo'lgan qiziqishini va yuqori natijalariga erishishdagi muvaffaqiyatini ta'minlaydi. Sun'iy intellekt vositalari faqatgina texnik yordamchi bo'lib, ular ijtimoiy va pedagogik kontekstda haqiqiy ta'limning o'rnini to'ldira olmaydi.

Inson omili va ta'limiy samaradorlik. Til o'rganish — bu faqat grammatika va lug'atlarni o'zlashtirish emas, balki madaniy, ijtimoiy va hissiy omillarni ham o'z ichiga olgan murakkab jarayondir. Sun'iy intellekt platformalari, masalan, Duolingo, ChatGPT yoki Replika kabi vositalar, albatta, tilni o'rganishda samarali yondashuvlarni taklif etadi. Biroq, bu texnologiyalar faqat kognitiv qobiliyatlarni rivojlantirishga qaratilgan va insonning ijtimoiy-emotsional, hamda madaniy jihatlarini to'liq hisobga olishda cheklangan. Ushbu vositalar foydalanuvchiga interaktiv mashqlar, grammatika tahlillari va avtomatik javoblar orqali bilimlarni mustahkamlash imkonini beradi, lekin ular insonning o'rganish jarayonidagi shaxsiy ehtiyojlari, hissiyotlari va madaniy o'ziga xosliklarni ko'rib chiqolmaydi.

O'qituvchi esa o'rganish jarayonida o'quvchiga faqat bilimlarni yetkazuvchi emas, balki psixologik yordamchi, ijtimoiy rahbar va tanqidiy fikrlashni rivojlantirishga yordam beruvchi muhim shaxs sifatida muhim rol o'ynaydi. O'qituvchining roli, ayniqsa, talaba o'z-o'zini anglash, ehtiyojlarni tushunish va o'z fikrlarini ifoda etishda yordam berishda katta ahamiyatga ega. O'qituvchi har bir o'quvchining individual xususiyatlarini inobatga olgan holda, ularni rag'batlantirishi, ichki motivatsiyasini oshirishi, shuningdek, eng yaxshi natijalarga erishish uchun zarur bo'lgan psixologik yordamni taqdim etadi. Bu esa, o'z navbatida, o'quvchining tilni o'zlashtirish jarayonida nafaqat axborotni, balki kommunikativ va ijtimoiy qobiliyatlarni ham rivojlantiradi.

Sotsiomadaniy nazariyaga ko'ra, ta'lim ijtimoiy muloqot orqali amalga oshadi. Bu nazariyaga ko'ra, o'rganish jarayoni ijtimoiy muloqotda mustahkamlanadi, ya'ni o'quvchilar o'zaro tajriba almashish, o'z fikrlarini bildirish va bir-birini tushunish orqali bilimni yanada chuqurlashtiradilar. Shuningdek, mazkur nazariyada shaxsning yaqin rivojlanish zonasini tushuntirib, o'quvchilar o'zining hozirgi bilim darajasidan biroz yuqoridagi, ammo hali erishilmagan darajaga o'qituvchi yordamida o'tishlari kerakligini ta'kidlagan. Bu jarayon esa, sun'iy intellekt tomonidan to'liq ta'minlanishi qiyin bo'lgan ijtimoiy va hissiy o'zaro aloqalarni o'z ichiga oladi.

SI vositalari bu muloqotni to'liq ta'minlay olmagani uchun, inson omilini ta'lim jarayonidan chetda qoldirish mumkin emas. Texnologiyalar, albatta, ta'limni samarali qilishda

yordam berishi mumkin, ammo ular insonning ijtimoiy muhit, madaniy differensial va hissiy xususiyatlarini inobatga olishda cheklovlarga ega. Sun'iy intellekt til o'rgatishda qimmatli vosita bo'lsada, uni faqatgina o'quvchilarning shaxsiy ehtiyojlarini va ijtimoiy o'rganish jarayonini to'liq inobatga olib ishlatish lozim.

Xulosa va tavsiyalar. Sun'iy intellekt vositalari tilni o'rganishda juda qulay va samarali vosita bo'lishi mumkin. Ular, xususan, mustaqil o'rganish uchun yo'l-yo'riqlarni avtomatlashtirib, har bir foydalanuvchining ehtiyojlariga moslashadi. Ushbu platformalar tomonidan taklif qilingan individuallashtirilgan va tezkor o'rganish jarayoni, zamonaviy talablarga javob beradi. Shu bilan birga, ularning afzalligi shundaki, foydalanuvchi o'z tempida va qulay vaqtda o'rganishi mumkin, bu esa motivatsiyani oshirishga xizmat qiladi. Ammo, bu vositalarning aniq bir chegaralari ham mavjud. Ularning ko'pchiligi, afsuski, chuqur ta'limiy yondashuvni, xususan, ijtimoiy, kommunikativ va madaniy kompetensiyalar rivojini ta'minlashda yetarli emas. Til o'rganish faqat nutq, grammatik bilim va so'z baytini shakllantirishdan iborat emas, balki ijtimoiy axborotni uzatish, madaniy muhitni tushunish, empatiya va muloqotni rivojlantirishni ham o'z ichiga oladi. Bu jihatlar esa faqat sun'iy intellekt vositalari bilan ta'minlanmaydi.

Eng samarali yondashuv, tadqiqot asosida kelib chiqqan xulosaga ko'ra, bu — **gibrid model**dir. Ya'ni, an'anaviy ta'lim usullari bilan sun'iy intellekt asosidagi vositalarni birgalikda qo'llash. Bu, o'qituvchilarga sun'iy intellekt platformalaridan yordamchi resurs sifatida foydalanish imkonini beradi, lekin ularning rolini pastga tushirishga yo'l qo'ymasligi kerak. Oqituvchilar jonli muloqotni saqlab qolishlari, talaba bilan to'g'ridan-to'g'ri ishlashlari va uning faol ishtirokini ta'minlashlari lozim.

Shunday qilib, kelajak ta'limi texnologiya va inson o'rtasidagi samarali hamkorlikka asoslanadi. Bu jarayonda insonning qadr-qimmati, madaniy xabardorlik va jonli muloqot har doim ustuvor o'rin tutadi. Ta'lim faqat ilmiy bilimlarni o'rgatish emas, balki insonning ijtimoiy, madaniy va psixologiyaviy jihatlarini rivojlantirishga qaratilgan bo'lishi kerak.

Adabiyotlar:

1. Azizova M. — "Zamonaviy ta'limda sun'iy intellektning o'rni", Toshkent, 2023.
2. Jumaev Sh. — "Lingvodidaktik yondoshuvlar va raqamli texnologiyalar", Samarqand, 2022.
3. Рейндэрс Х. — "Цифровое обучение и преподавание языков", Palgrave Macmillan, 2020. DOI:10.1007/978-3-030-46244-2
4. Лю, М. — "AI-powered Language Learning Apps: Opportunities and Challenges", *Language Learning & Technology*, 2021.
5. Turobova N. — "Til o'rganishda psixologik omillar", Namangan, 2023.
6. Duolingo Research Report, 2022. [<https://research.duolingo.com>]
7. OpenAI: ChatGPT in Education, 2023. [<https://openai.com/education>]
8. Replika Research on Communication Skills, 2022. [<https://replika.ai/research>]

TABIIY TIL MATNLARINI AVTOMATIK GRAMMATIK VA ORFOGRAFIK TAHRIRLASHDA ALGORITMIK YONDASHUVLAR TAHLILI

Nazirova Elmira Shodmonovna

Texnika fanlari doktori, professor

Muhammad al-Xorazmiy nomidagi

Toshkent axborot texnologiyalari universiteti

Boymurodov Farrux Farxod o'g'li

Doktorant, Qarshi davlat universiteti

Annotatsiya: Hozirgi kunda Internetda matnli axborot hajmining keskin ortib borayotganligi, tabiiy tildagi matnlarni (NLP-matnlar) avtomatik qayta ishlovchi dasturiy ta'minot tizimlarini rivojlantirishda algoritmnlarni ishlab chiqish dolzarb ahamiyat kasb etmoqda. Bunday tizimli algoritmnlarni ishlab chiqish jarayonida bir qator muammolar yuzaga keladi. Jumladan, matnni so'z, gap yoki ibora darajasida qayta ishlovchi ko'plab dasturiy algoritmnlarni yagona tizimga birlashtirish zarurati mavjudligidir. Ushbu tezisda tabiiy til matnlarini tahlil qilishga yaratilgan alohida quyi tizimlar uchun zarur bo'lgan model va algoritmnlar bayon etiladi. Matnli ma'lumotlarni tasniflash va klasterlash tizimlarini yaratish masalalari hamda matnli axborotni fraktal nazariya asosida o'rganish tamoyillari ko'rib chiqiladi. Shuningdek, asosiy lingvistik obyektlar va tahlillar talqinlariga alohida e'tibor qaratiladi.

Kalit so'zlar: NLP, imlo tahrir, gramatik tahrir, Bayes, Markov zanjiri, UZMORF.

Аннотация: В настоящее время наблюдается стремительный рост информационных ресурсов в Интернете, разработка программного обеспечения для обработки естественного языка (NLP) для автоматической обработки информационных ресурсов, разработка алгоритмов обработки текста, а также развитие профессиональных навыков в области обработки текста. В процессе разработки таких системных алгоритмов возникает ряд проблем. В частности, необходимость интеграции множества программных алгоритмов для обработки текста на уровне слова или фразы в единую систему. В диссертации описываются модели и алгоритмы, необходимые для конкретных систем, способных анализировать тексты на естественном языке. Рассматриваются вопросы систем классификации и кластеризации текстовой информации, а также результаты изучения текстовой информации на основе теории фракталов. Особое внимание уделяется интерпретации своевременных, базовых и надежных анализов объектов.

Ключевые слова: NLP, орфографическая коррекция, грамматическая правка, Байес, цепь Маркова, UZMORF.

Annotation: Nowadays, the rapid growth of information resources on the Internet, the development of natural language processing (NLP) software for automatic processing of information resources, the development of algorithms for text processing, and the development of professional skills in the field of text processing. In the process of developing such systematic algorithms, a number of problems arise. In particular, the need to integrate many software algorithms for processing text at the word or phrase level into a single system. The thesis describes the models and algorithms necessary for specific systems capable of analyzing natural language texts. The issues of classification and clustering systems for text information, as well as the results of studying text information based on fractal theory, are considered. Special attention is paid to the interpretation of timely, basic and reliable object analyses.

Keywords: NLP, spelling correction, grammatical editing, Bayes, Markov chain, UZMORF.

Tabiiy til matnini yuqori sifatga ega qilish uchun uni qayta ishlash jarayonlarida to'liq va mukammal tahlilni amalga oshirish zarur. Ammo, bunday tahlilni yaratishda duch kelinadigan asosiy qiyinchiliklar – matnni tahlil qilishning murakkabligi va o'zbek tiliga xos barcha qoidalarini qamrab oluvchi tizimni yaratishning qiyinligidir. Ushbu muammolarni bartaraf etish uchun matnni sintaktik (gap tuzilishi), semantik (mazmun va ma'no) va pragmatik (tushunchalarning to'g'ri qo'llanilishi) jihatdan tahlil qilish imkoniyatiga ega bo'lgan tizim zarur [Bolshakova E.I, 2011; 31].

O'zbek tilidagi matnlarni grammatik va orfografik jihatdan avtomatik tahrirlash masalasi tabiiy tilni qayta ishlash (NLP) sohasida dolzarb yo'nalishlardan biri hisoblanadi. Bu masalaning murakkabligi o'zbek tilining agglutinatativ xususiyatlari, morfologik o'zgarishlari va sintaktik strukturalarni hisobga olishi talab etiladi. Ushbu tadqiqotning asosiy maqsadi – o'zbek tilidagi matnlarni grammatik va orfografik jihatdan avtomatik tahlil qilish va xatolarni to'g'irlash imkoniyatiga ega bo'lgan dasturiy tizim yaratishdir. Tizim Microsoft Word muhiti bilan integratsiyalashuvi amalga oshiriladi va foydalanuvchi tomonidan kiritilgan matnni tahlil qiladi va xatolarni to'g'irlaydi [Boymurodov F, 2025; 39].

O'zbek tilidagi matnlarni grammatik va orfografik jihatdan avtomatik tahrirlash jarayonini quyidagi boshlang'ich model asosida ko'rib chiqiladi. Matnning dastlabki shakli, undagi harflar, qo'shimchalar va so'z shakllari noto'g'ri yozilgan bo'lishi mumkin. Natijada, foydalanuvchi tomonidan kiritilgan matn $\hat{C}$ ko'rinishida, ya'ni xatolar bilan ifodalanishi mumkin. Xatolarni avtomatik tuzatish uchun ehtimollik nazariyasi, **Bayes** klassifikatori va **Markov zanjiri** asosida quyidagi model va algoritmlar taklif etiladi:

$$P(C | \hat{C}) = \frac{P(\hat{C} | C)P(C)}{P(\hat{C})}$$

bu yerda:

$P(C | \hat{C})$ – xato yozilgan $\hat{C}$ so'zining to'g'ri C so'z bo'lish ehtimoli,

$P(\hat{C} | C)$ – C so'zining xato yozilgan muqobil variantlari,

$P(C)$ – so'zning doimiy ishlatilish holati,

$P(\hat{C})$ – kiritilgan noto'g'ri so'zning paydo bo'lish ehtimoli.

Markov zanjiri. Ushbu algoritm tasodifiy jarayon bo'lib, uning asosiy xususiyati shundaki, **kelajak faqat hozirgi holatga bog'liq bo'ladi** va oldingi holatlarga bog'liq emas. Matnni qayta ishlashda Markov modeli har bir so'z yoki harfning **oldingi so'z yoki harflar asosida quyidagi ketma-ketlikda paydo bo'lish ehtimolini** hisoblaydi.

- matn katta bazadan tekshiriladi. Har bir so'z yoki harfning keyingi so'z yoki harf bilan bog'lanish ehtimolligi hisoblanadi.

- agar noto'g'ri so'z paydo bo'lsa, eng ehtimolli keyingi so'z bilan taqqoslanadi.

- eng mos keluvchi to'g'ri variant tanlanadi va o'sha so'z bilan almashtiriladi.

Faraz qilaylik, bizda ushbu matn bor: **“Men bugun kitb o'qiyapman.”**

Markov modeli quyidagi tarzda ishlaydi:

- oldingi matn bazasidan **“kitob”** so'zi ko'p uchrayotganini tahlil qiladi.

- **“kitb”** dan keyin **“o'qiyapman”** kelishi **“kitob”** bilan mos kelishini aniqlaydi.

- eng ehtimolli variant – **“kitob”** bo'lgani uchun, noto'g'ri yozilgan so'z almashtiriladi.

Bayes klassifikatori mustaqil ehtimollik asosida so'zlarni tahlil qiladi va xatolarni aniqlash uchun tezkor yechim beradi.

Markov zanjiri esa so'zlar orasidagi bog'lanishni hisobga oladi va kontekst asosida xatolarni to'g'ri variant bilan almashtiradi.

1-jadval. Model xususiyatlari

Xususiyat	Bayes klassifikatori	Markov zanjiri
Ishlash tamoyili	Mustaqil ehtimollar bo'yicha tahlil qiladi	So'zlarning ketma-ketligini hisobga oladi

Kontekstni hisobga oladimi?	Hisobga olmaydi, soʻzlarni mustaqil baholaydi	Hisobga oladi, oldingi soʻzlar taʼsir qiladi
Qanday xatolarni topadi?	Xatolarni ehtimollar boʻyicha aniqlaydi	Xatolarni ketma-ketlik asosida topadi
Xatolarni toʻgʻirlash yondashuvi	Toʻgʻri soʻzni tanlash orqali	Matn ketma-ketligidan kelib chiqib
Amalda ishlatilishi	Imlo tekshirish, gramatik tekshirish	Matnni generatsiya qilish
Afzalligi	Katta matnlar bilan tez ishlaydi	Kontekstga asoslangan natijalar beradi
Kamchiligi	Kontekstni inobatga olmaydi	Katta matnlarni tahrirlash koʻp vaqt talab etadi

Ushbu model algoritmlari orqali oʻzbek tili matnlarini avtomatik tahrirlash algoritmlarini ishlab chiqish va joriy etish oldimizga qoʻyilgan masala hisoblanadi. Tahlil qilingan algoritmlarni joriy etish uchun UZMORF nomli dastur pluginlarini ishlab chiqish joriy etiladi. Kelajakda UZMORF dasturi oʻzbek tilidagi matnlarni avtomatik ravishda qayta ishlash uchun ishlab chiqilgan dasturiy vosita hisoblanadi. Bu dastur oʻzbek tilidagi soʻzlarning morfologik tuzilishini tahlil qilib, ularni kompyuter orqali qayta ishlash uchun moslashtiradi. UZMORF dasturining ishlash prinsiplari quyidagicha amalga oshiriladi:

Soʻzlar va ularning morfologik tuzilishi: UZMORF dasturi oʻzbek tilidagi har bir soʻzning morfologik tuzilishini aniqlashga qaratilgan, yaʼni bu jarayonda soʻzning turkumi, yasallish shakllari va boshqa morfologik xususiyatlarini aniqlashga qaratiladi.

Morfologik tahlil: dastur oʻzbek tili morfologiyasining qoidalarini va xususiyatlariga asoslanadi. Matndagi har bir soʻz tahlil qilinib, uning morfologik maʼlumotlari aniqlanadi. Bu jarayon morfologik analiz algoritmlari yordamida amalga oshiriladi.

Yordamchi dasturlar va modullar: UZMORF Python dasturlash tilida ishlab chiqiladi, matnlarni oʻzbek tilidagi soʻzlar va ularning tuzilishi boʻyicha tahlil qilish uchun maxsus modullardan foydalaniladi. Ushbu modullar matnlarni qayta ishlash va morfologik tahlillarni tezkorlik bilan amalga oshirish imkonini beradi.

Oʻzbek tiliga xos maʼlumotlar bazasi: dastur oʻz ichiga oʻzbek tiliga xos soʻzlar va ularning morfologik xususiyatlarini jamlagan maʼlumotlar bazasini oladi. Bu baza yordamida dastur matnlarni tahlil qilishda aniq va ishonchli natijalar berishi koʻzda tutilgan.

Yangilanish va texnik taraqqiyotlar: UZMORF dasturi zamonaviy texnologiyalarni hisobga olgan holda yangilanib boradi. Unda yangi oʻzbek tilidagi soʻzlar, ularning morfologik tuzilishi va tahlil qilish usullari muntazam qoʻshilib, dastur imkoniyatlari kengaytirib borish xususiyatlariga ega boʻladi.

Ushbu usullarni birgalikda qoʻllash matnlarning grammatik va stilistik jihatdan toʻgʻriligini taʼminlaydi. Oʻzbek tilidagi matnlarni tahrirlash uchun ishlab chiqilgan tizimlar, ushbu usullarga asoslanib, soʻzlarni avtomatik tahlil qilish va tuzatish imkoniyatini beradi. Yaratilayotgan dastur Microsoft Word muhiti bilan integratsiyasi amalga oshiriladi, u foydalanuvchilar uchun maksimal qulaylikni taʼminlaydi va oʻzbek tilidagi matnlarni tahrirlash jarayonlarini sezilarli darajada soddalashtiradi. Bu kabi loyihalarni kengaytirish va rivojlantirish oʻzbek tilining raqamli texnologiyalar sohasida keng qoʻllanishini taʼminlaydi.

Foydalanilgan adabiyotlar:

1. Эшкароева Н.Г., Боймуродов Ф.Ф. Разработка и реализация алгоритма стемматизации и генерации словоформ для глаголов узбекского языка //ҚарДУ ХАБ. – С. 4.

2. Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / — М.: МИЭМ, 2011
3. Эшкараева Н. Г., Боймуродов Ф. Ф. Разработка и реализация алгоритма стемматизации и генерации словоформ для глаголов узбекского языка //ҚарДУ ХАБ. – С. 4.
4. Nazirova E, Uzakova M, Boymurodov F Turkiy tillar oilasiga mansub o ‘zbek va qirg ‘iz tillarida elektron tarjimaning lingvo-matematik modelini ishlab chiqish: Turkiy tillar oilasiga mansub o ‘zbek va qirg ‘iz tillarida elektron tarjimaning lingvo-matematik modelini ishlab chiqish //MODERN PROBLEMS AND PROSPECTS OF APPLIED MATHEMATICS. – 2024. – T. 1. – №. 01.
5. Boymurodov F. MICROSOFT WORD UCHUN O ‘ZBEK TILI SO ‘ZLARNI IMLOVIY TAHRIRLASH PLAGININI YARATISH ALGORITMLARI VA DASTURI //DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE. – 2025. – T. 3. – №. 1. – С. 39-44.
6. Boymurodov F. ALGORITHM DEVELOPMENT AND IMPLEMENTATION STEMMATIZATION AND GENERATION OF WORD FORMS OF UZBEK LANGUAGE FOR AUTOMATIC PROCESSING SYSTEMS TEXTS //Конференции. – 2020.

NUTQ VA USLUB

Taxirov Z.T.

O'zMU jurnalistika va filologiya fakulteti professori,
filologiya fanlari doktori

Annotasiya: Mazkur maqolada amaliyotda to'g'ri ekanligi isbotlangan vabir necha talablarga asoslangan holda nutq va uslub tahlil etilgan.

Tayanch termin va atamalar: Nutq, uslub, amaliy uslubiyat, metodika, o'xshatish, matn baholash, sintaktik konstruksiya, tanqidiy fikrlash, muallif matni, ortiqcha so'z, odatiy xato.

Аннотация: В данной статье представлен анализ речи и стиля, основанный на нескольких требованиях, правильность которых доказана на практике.

Annotation: This article presents an analysis of speech and style based on several requirements that have been proven correct in practice.

Ma'lumki, "... nutq va uslub tahlili, umuman aytganda, adabiy tahrir va amaliy uslubiyat bilan bog'liq"[1].

Har ikki fan ham tahlilning amaldagi metodikasini ifoda etmaydi. Nutq va uslub ega bilan kesimni gapda qanday moslashtirishni, bir xil yoki bir o'zakli so'zlar muayyan maqsadda qo'llanmagan bo'lsa, to'g'ri emasligini, imloviy xato qilmaslik, tinish belgilaridan to'g'ri foydalanish va b.ni o'rgatadi. Bu fanlar tufayli jumla tuzish, gap bo'laklarining joylashishi, bir-biri bilan bog'lanishini, so'zlarni o'z denotativ ma'nosiga ko'ra yoki kennotativ ma'noda qo'llanishi va h.lar ayonlashadi.

Amaliyotda to'g'riligi isbotlangan bir necha talablarga amal qilgan holda matn bayoni (nutq) va uslubni tahlil etish jiddiy xatolarning oldini oladi. Ta'rif va tushuntirishdan iborat nutq "...juda oddiy (tuzilishi nuqtai nazardan), ancha barqaror va mantiqan qat'iylashgan nutq"[2] hisoblanadi.

Matnni til va uslubga ko'ra muvaffaqiyatli tahlil qilishga qo'yiladigan talablar quyidagicha:

1) matnni tahlil qilishni til va uslubga ko'ra umumiy va o'ziga xosligini aniqlashdan boshlash kerak.

2) muallifning so'zlardan foydalanishligi va sintaktik konstruksiyalar tuzish tarzini aniqlash lozim.

Agar muallif matnda tavgologiyaga, qaytariq iboralarga yo'l qo'ygan bo'lsa, ya'ni:

– *Fanlar akademiyasining fanlarga oid tadqiqotlarni e'lon qilishi gumanitar va tabiiy fanlar uchun qanday ahamiyatga egaligi...*

Ko'rinadiki, muallif bitta jumla tarkibida bir o'zakli, bir xil so'zlar (*fanlar, fanlarga, fanlar*)ni qo'llagan.

Matnning til va uslubga ko'ra o'ziga xos jihatlarni aniqlash, uni aniq tasavvur etish uchun zamin yaratadi.

Matndagi qaytariqlar (*aytaylik, shunday qilib, alohida ahamiyat kasb etadi* va b.) go'yo matnda rang-baranglikni ta'minlashga xizmat qilsa-da, lekin uslub g'alizligiga sabab bo'ladi.

Matnni til va uslub bo'yicha o'ziga xosligini aniqlash, bu – hali etarli emas. Asosiysi, matnni til va uslubga ko'ra maxsus tahlil qilishdir. Chunki maxsus tahlilsiz matn muallifining individual (ayniqsa, badiiy matndagi) uslubini mukammal yoki mukammal emasligini aniqlash deyarli mumkin bo'lmaydi.

Agar tasvir va hikoya uchun kesim shaklini tanlashda sintaktik paralellik [jc bo'lsa, (bu jumlar tuzilishida ko'rinadi), "mulohazali nutq"ning asosiy birlamchi nasriy satr zanjir halqalari singari bog'lanadi" [3].

Badiiy matn ijodchilari badiiylilik, obrazlik va ta'sirchanlikni oshirish maqsadida g'ayritabiiy o'xshatish qiladilar, bu, albatta, maxsus tahlil qilishni talab etadi.

Oybekning Navoiy romanini o'qigan kishi, adibning o'xshatishlari naqadar o'ziga xos ekanligiga guvoh bo'ladi.

Dasturxon yozildi. Kosalar, piyolalar, qadahlarning gullari bahor chamanlariday boy, nafis, rangin edi.

Mazkur parcha til va uslubga ko'ra tahlil etilishi uchun uning tarkibidagi unsurlar (qora harflardagi) alohida-alohida tahlil qilinishi lozim:

... *piyolalar, qadahlarning gullari bahor chamanlariday.*

Bunda o'xshatish morfologik vosita – *day* yordamiga yuzaga chiqqan.

Albatta, matn tahlili mantiqqa asoslanadi. Til, uslub, mantiq doirasidagi tuzatishlar o'rinli bo'lishi mumkin.

Shuni alohida ta'kidlash joizki, tasviriy fanlar bilan uslubiyatning tadqiqot obyekti ayni bir til hisoblanadi. Biroq, o'rganish predmeti bosqqa-boshqa. Til tizimi til tizimidagi birliklar va kategoriyalar tasviriy fanlarning predmetini tashkil etadi. "Tildan, ayniqsa, adabiy nutqdan foydalanishning barcha usullari uslubiyatning predmeti hisoblanadi"[4].

Narsa, voqea, hodisaga bo'lgan muallif ruhiy munosabati nutq me'yorlari doirasidan chetga chiqib ketishi mumkin.

Matnni o'quvchi, agar nutq egasining ruhiyatiga kira olsa, unda go'yo xato tuyulayotgan ifoda, iboradan muallifona tushuncha, tuyg'uni ilg'ay oladi.

Nutqni tahlil qilishda, ya'ni unga uslubiy jihatdan tuzatishlar kiritish, xususida tavsiyalar berishda sub'ektivlikka yo'l qo'ymaslik kerak.

Uslubga oid ko'pgina tuzatishlarni ma'qullash mumkin, muharrirning intilishi ham tushunarli. Chunki, har bir tahrirchi matnni ayni o'zi uchun tushunarli holga keltirishga bo'lgan xohishiga bo'ysunadi. Tahrir shunday bo'lishi kerak, deb biladi.

Matndagi ayrim mantiqiy-sintaktik konstruksiyalar murakkabligi, nashr talabiga mos emasligiga ko'ra o'zgartirilishi kerak. O'quvchi matnni o'qir ekan ko'pincha uni yaxshi tushunish, yana ham aniqrog'i, eslab qolish uchun xayolan qayta tuzadi. Chunki, matn qancha qisqa va umumlashtirilgan bo'lsa, nutqdan anglashiladigan mazmunning o'zlashtirilishi osonlashadi.

Matnni fikran qayta tuzish va o'z fikrlash tartibiga moslash xatoni to'g'rilash, degani emas. Ayrim tahrirchilar nutq egasining ifodasini qabul qilolmaydi. Masalan:

Asta-sekin – yozadi muallif. *Sekin-asta* – tuzatadi muharrir.

Demak – yozadi muallif. *Shunday qilib* – tuzatadi muharrir.

Vaholanki – yozadi muallif. *Holbuki* – tuzatadi muharrir.

Xatoga yo'l qo'yishga, odatda, yozishdan ko'ra fikrlashning tezligi sabab bo'ladi – yozadi muallif.

Odatda, fikrlashning yozishdan ko'ra tezligi xatoga yo'l qo'yilishiga sababdir – o'zgartiradi muharrir.

Muallif va tahrirchining fikrlash uslubi, nutq tarzi turli bo'lishi mumkin.

Tahrirchi muallifning uslubini emas, balki, uslubdan noto'g'ri foydalanganini aniqlashi kerak.

Matnga subyektiv yondashish, uni o'z fikrlash tizimiga moslashga bo'lgan intilishlardan holi bo'lish uchun til me'yorlarini yaxshi bilish kerak. Tahlilda kontekstning o'ziga xosligi va vazifasini to'g'ri idrok etish muxim ahamiyat kasb etadi.

Matn til va uslubga ko'ra tahlil etilar ekan, me'yoriy uslubiyat qoidalari hamda tavsiyalaridan oqilona, foydalana olishi talab etiladi. Noshirlik sohasida faoliyat ko'rsatayotgan muharrirlar (OAV tahririyatidagilar ham) bevosita o'zlari tahrirdan chiqarayotgan, chiqarishi

kerak bo‘lgan matnga uslubiyatning qaysi me‘yorlariga ko‘ra yondashishlari lozimligini yaxshi idrok etbilari kerak.

Yozma nutq, xususan, matn ifodasi nihoyatda nozik masala, unga bir qolipda yondashish asarni nutqiy, uslubiy jihatdan yaxshilash o‘rniga, uni abgor qilishi, tayin.

“Erto‘g‘ril” filmidagi din arbobining nutqiga e‘tibor beraylik.

Tangrining nuri hamisha salibchilar yo‘lini yorug‘ qilsin.

Tarjimon filmdagi personajlarni “sof” o‘zbek tilida so‘zlatmoqchi bo‘lgandir. *Tangri* – turkiylarni yakka ilohi. Salibchilar katolik, xristianlar – xochparast, mushriklar. Bunday nutqiy qashshoqlikka, tilni bilmaslik, tarjimon o‘z bilganicha ish ko‘rganligi, subyektivlikka yo‘l qo‘yganligi sabab bo‘lgan. Tarjimon ham, tahrirchi ham turkiylar yakka ilohi *Tangrini* mushriklarniki qilib qo‘ygan. So‘z qo‘llashda o‘ta qo‘pol xatoga yo‘l qo‘yilgan.

“Tilni bilish” faqat amalda iboralar, sintaksis me‘yorlaridan bilimga ega bo‘lish, degani emas, balki til va uslubni nazariy jihatdan mukamal egallagan bo‘lish zarur. Chunki amaliyotda til va uslub bilan bog‘liq ko‘plab vazifani bajarishga to‘g‘ri keladi. Buni esa faqat chuqur nazariy bilimga tayanibgina uddalash mumkin. Lekin, til va uslubga oid xatolarni aniqlashga bu hali yetarli emas. “Uslubiyat nutq madaniyatidan farqli o‘laroq asosiy e‘tiborni adabiy nutqga qaratadi”[5].

Xullas, matnni bunday xatolardan holi qilish uchun tahlil va tahrir nazariyasi hamda metadologiyasini a‘lo darajada bilish kerak.

Adabiyotlar ro‘yati:

1. Taxirov Z.T. Tahrir mantiqi. –T. O‘zbekiston, 2016. –187-b.
2. Одинцов В.В. Стилистика текста. –Москва, 1980.– 381-b.
3. Солганик Г.Я. Синтаксическая стилистика. –Москва, 1991. –7–8-b.
4. Владимиров В.В. Проблемы русской стилистики. –Москва, 1981. –173-b.
5. Taxirov Z.T. O‘zbek tilining amaliy uslubiyati. –T.: Tafakkur avlodi, 2021. –18-b.

YOZMA TARJIMA KO'NIKMASINI AVTOMATLASHTIRILGAN TARJIMA VA LINGVISTIK TAHLIL VOSITALARI YORDAMIDA OSHIRISH

Axmedova Anorxon Nasivali qizi

O'zDJTU v.b. dotsenti,
filologiya fanlari bo'yicha falsafa doktori (PhD)

Zamonaviy tarjimashunoslikning jadal rivojlanishi, madaniy aloqalarning kengayib borishi va global integratsiya jarayonlari natijasida tarjimonlarning, ayniqsa, yozma tarjima sohasida faoliyat yuritayotgan mutaxassislarning roli yanada ortib bormoqda. Xalqaro miqyosda badiiy adabiyotlarning tarjima qilinishi nafaqat adabiy almashinuvni, balki turli xalqlar o'rtasida madaniy hamkorlik va o'zaro anglashuvni mustahkamlashda muhim vosita bo'lib xizmat qilmoqda. Shu bois, tarjima jarayonining sifatini oshirish, uni zamonaviy metodlar asosida o'qitish va kelajak tarjimonlarining kasbiy tayyorgarligini takomillashtirish bugungi kunda dolzarb masalaga aylanmoqda.

Tarjima faoliyatida til birliklarining nozik ma'no farqlari, konnotativ ma'no, madaniy realiyalar va badiiy ifoda vositalarining to'g'ri talqin qilinishi katta ahamiyatga ega. Bunday murakkab lingvistik birliklarni chuqur tahlil qilish va ularning ekvivalentlarini aniqlashda korpus lingvistikasi zamonaviy vosita sifatida keng qo'llanilmoqda. Ayniqsa, korpusga asoslangan yondashuv tarjima jarayonini o'qitishda, tarjimonlik malakalarini rivojlantirishda va tilshunoslik nuqtayi nazaridan mustahkam tahliliy bazani yaratishda o'zining samaradorligini isbotlamoqda.

Korpus lingvistikasi asosida yaratilgan ikki va ko'p tilli korpuslar tarjima ta'limida real va kontekstual foydalanish imkoniyatini beradi. Talabalar korpusdan foydalanish orqali asl va tarjima matnlaridagi til birliklarini taqqoslaydi, tarjimada ishlatilgan strategiyalarni o'rganadi, tarjimaning adekvatligi va ekvivalentligini tahlil qilish imkoniga ega bo'ladi. Bu esa, o'z navbatida, bo'lajak tarjimonlarning analitik fikrlash qobiliyatini, lingvistik sezgirligini va ijodiy yondashuvini rivojlantirishga xizmat qiladi.

Shu jihatdan qaraganda, korpusga asoslangan tarjima o'qitish metodikasini ishlab chiqish, uni amaliyotga joriy etish va bu borada samarali pedagogik modelni taklif qilish bugungi kunda tarjimashunoslik va til o'qitish sohasining dolzarb va istiqbolli yo'nalishlaridan biri hisoblanadi.

So'nggi yillarda yozma tarjima bo'yicha ta'lim jarayoniga korpus lingvistikasi asosida yondashuvlar jadal kirib kelmoqda. Til korpuslari – bu real nutq namunalari asosida tuzilgan raqamli matnlar bazasi bo'lib, tarjimonlik faoliyatida til birliklarining kontekstda qo'llanishini kuzatish, tarjima ekvivalentlarini aniqlash va tahlil qilishda juda katta imkoniyat yaratadi. Ayniqsa, ikki tilli va ko'p tilli korpuslar asosida tarjimoni o'rgatish talabalarning tahliliy fikrlash, so'z va ifodalarni to'g'ri kontekstdan kelib chiqib tanlash, tarjima strategiyalarini asoslash, badiiy-estetik mezonlarga e'tibor qaratish kabi ko'nikmalarini shakllantiradi.

Yozma tarjimoni til korpuslari asosida o'qitish jarayonida matnlarning lingvistik, madaniy va uslubiy qatlamlarini tahlil qilish imkoniyati kengayadi. Talabalar tarjima qilingan matnlarni korpusda tahlil qilish orqali sinonimik qatlam, konnotatsiya, madaniy realiyalar, stilistik vositalar va boshqa tarjima muammolarini amaliy misollar orqali chuqur anglay boshlaydilar. Bu esa ularning umumiy, lingvistik, kasbiy, ijodiy va ensiklopedik kompetensiyalarini bosqichma-bosqich rivojlantirishga xizmat qiladi.

Shu jihatdan, yozma tarjimoni til korpuslari asosida o'qitish – nafaqat tarjimonlikni o'rgatishning zamonaviy va samarali metodikasi, balki u orqali talabalarning tarjimaga yondashuvi, tanqidiy tahlil qilish qobiliyati, mustaqil o'rganish va texnologiyalardan foydalanish ko'nikmalarini shakllantirishda muhim vosita hisoblanadi.

Bugungi kunda tarjimonlar uchun qulay shart-sharoitlar yaratadigan ko'plab zamonaviy vositalar mavjud bo'lib ular tarjimonlarning eng yaqin yordamchisi sifatida xizmat qilib kelmoqda. Xususan, yozma tarjima ko'nikmasini rivojlantirishda avtomatlashtirilgan tarjima tizimlari

(masalan, Google Translate, DeepL, Reverso) va lingvistik tahlil dasturlarining oʻrni juda katta. Bu vositalar tarjimonlarga nafaqat soʻzma-soʻz tarjima qilishda, balki matnning umumiy mazmunini tushunish, grammatik qurilmalarning tuzilishini tahlil qilish va stilistik jihatdan mos ekvivalentlarni tanlashda ham katta yordam beradi. Masalan, Reverso Context yordamida biror iboraning kontekstda qanday ishlatilishini koʻrib chiqish mumkin. Bu esa tarjimonning madaniy va uslubiy nozik jihatlarga eʼtibor berib tarjima qilishiga yordam beradi. Shuningdek, lingvistik tahlil dasturlari (masalan, Sketch Engine, AntConc) orqali matndagi asosiy soʻz birliklari, kollokatsiyalar, va leksik model tahlil qilinadi. Bu tarjimada anʼanaviy iboralarni toʻgʻri tanlash, sinonimlarni ajratish va matn uslubini saqlash imkonini beradi.

Yozma tarjimani oʻqitishda zamonaviy yondashuvlar orasida texnologiyalardan foydalanish alohida oʻringa ega. Zamonaviy oʻqitish jarayonida avtomatlashtirilgan koʻplab tarjima vositalari va lingvistik tahlil dasturlari foydalaniladi. Talabalar ushbu vositalar yordamida tarjimani tezkorlik bilan amalga oshirishni va natijalarni tahlil qilishni oʻrganadilar. Quyidagi jadvallarda avtomatlashtirilgan tarjima vositalari va lingvistik tahlil dasturlariga alohida toʻxtalib oʻtamiz.

1-jadval. Avtomatlashtirilgan tarjima vositalari

№	Avtomatlashtirilgan tarjima vositalari
1.	Microsoft Translator – matn, nutq va rasm tarjimalarini amalga oshiradi.
2.	Amazon Translate – AWS platformasiga asoslangan avtomatik tarjima xizmati.
3.	Systran – sunʼiy intellektga asoslangan ilgʻor tarjima tizimi.
4.	Smartcat – tarjimonlar, muharrirlar va mijozlar oʻrtasida hamkorlikni taʼminlovchi tarjima platformasi.
5.	MateCat – onlayn rejimda ishlovchi tarjima vositasi boʻlib, katta hajmdagi matnlarni tarjima qilishga imkon beradi.
6.	Yandex.Translate – ruscha koʻp tilli tarjima tizimi.
7.	Reverso Context – kontekstual tarjima boʻlib interfeysda izlangan soʻz yoki iboralarning tarjimasini sinonimlari bilan taqdim etish xususiyatiga ega tizim hisoblanadi.
8.	Papago (by Naver) – Janubiy Koreyaning ilgʻor tarjima tizimi boʻlib, ayniqsa sharqiy Osiyo tillari uchun foydali tizimdir.
9.	iTranslate – Mobil ilovalarga moʻljallangan ilgʻor tarjima tizimi.
10.	SayHi – asosan ogʻzaki tarjima uchun ishlatiladigan tizim.
11.	MemoQ – professional tarjimonlar uchun tarjima xotirasi tizimi.
12.	Wordfast – yengil, samarali va integratsiyalashgan tarjima xotirasi tizimi.
13.	CafeTran Espresso – tarjimonlar va tarjima agentliklari uchun yaxshi moslashuvchan tizim.
14.	Lokalise – ilovalar va veb-saytlarni tarjima qilish va lokalizatsiya qilish vositasi.
15.	Poedit – dasturiy taʼminot uchun tarjima qilish va lokalizatsiya vositasi.
16.	OmegaT – bepul va ochiq kodli tarjima xotirasi tizimi.
17.	Crowdin – ilovalar, oʻyinlar va veb-saytlarni tarjima qilish uchun platforma.

2-jadval. Lingvistik tahlil vositalari

№	Lingvistik tahlil dasturlari (Linguistic Analysis Tools)
1.	Grammarly – matn grammatikasini, uslubini va tushunarli boʻlishini tekshiradi.
2.	Hemingway Editor – matnning oʻqilishi va uslubiy aniqligini yaxshilash uchun ishlatiladi.
3.	AntConc – korpus lingvistikasi boʻyicha tahlil vositasi, soʻz chastotasi va kollokatsiyalarni tekshiradi.
4.	Voyant Tools – matn tahlili uchun vizualizatsiya va statistik maʼlumotlar taqdim etadi.
5.	Sketch Engine – lingvistik tadqiqotlar uchun korpus tahlili dasturi.

6.	ProWritingAid – grammatik tahlil, uslubni yaxshilash va takrorlanishlarni aniqlash imkoniyatiga ega.
7.	Quyidagi vositalar matn tahlili, grammatik tahlil, leksik tahlil va korpus lingvistikasi uchun ishlatiladi: 1. Grammatik va sintaktik tahlil vositalari LanguageTool – grammatik xatolarni avtomatik aniqlash va tuzatish. WhiteSmoke – yozuv uslubini yaxshilash va grammatik xatolarni tuzatish uchun ishlatiladi. QuillBot – matnni qayta yozish va tahlil qilish vositasi. Ludwig – ingliz tilidagi jummalarni kontekst bo'yicha tahlil qilish va to'g'ri tuzish uchun ishlatiladi.
8.	2. Korpus lingvistikasi va semantik tahlil vositalari UAM CorpusTool – korpus lingvistikasi tahlili va belgilar tizimini yaratish. GATE (General Architecture for Text Engineering) – matnni tahlil qilish, leksik tahlil qilish va AI asosida ishlaydigan NLP tizimi. Stanford NLP – lingvistik tahlil uchun kuchli dasturiy ta'minot, morfologik va sintaktik tahlilni bajaradi. TreeTagger – so'z turkumlarini aniqlash va lingvistik tahlil qilish.
9.	3. So'z chastotasi va semantik tahlil vositalari WordSmith Tools – so'z chastotasi va kollokatsiyalarni tahlil qilish uchun ishlatiladi. LIWC (Linguistic Inquiry and Word Count) – matnning psixologik va ijtimoiy jihatlarini tahlil qiladi. Textalyser – matnning lingvistik statistikasi, chastotasi va uzunligini hisoblash.
10.	4. Fonetik va talaffuz tahlili dasturlari Praat – Fonetik tahlil va nutqni o'rganish uchun eng mashhur dastur. ELAN – Nutq va video tahlili uchun lingvistik vosita. Phon – Fonologik va talaffuz tahlili uchun mo'ljallangan dastur.

Umuman olganda, ushbu texnologik vositalar orqali talaba yoki tarjimon o'z bilimini mustahkamlab boradi, tarjima xatolarini tahlil qiladi va asta-sekin o'z tarjima uslubini shakllantiradi. Yozma tarjimani o'qitishda korpus lingvistikasi va raqamli vositalardan foydalanish zamonaviy ta'limning talablariga to'la javob beradi. Bu metodika nafaqat tarjimonlar malakasini oshirishga, balki ularning ijodiy, lingvistik va tanqidiy fikrlash salohiyatini rivojlantirishga ham samarali hissa qo'shadi. Shu asosda kelajakda yanada mukammal, milliy va xalqaro korpuslar bilan ishlashga yo'naltirilgan kompleks pedagogik modellarni ishlab chiqish va amaliyotga joriy etish dolzarb ilmiy-amaliy vazifalardan biri bo'lib qolmoqda.

YURIDIK DISKURS MATNLARI

Mirvaliyeva Dildora

O'zbekiston Milliy universiteti o'qituvchisi

E-mail: dildoramirvaliyeva07@gmail.com

Annotatsiya. Maqolada diskurs, yuridik diskurs va yuridik matnlarning xususiyatlari haqida so'z boradi. Til va huquq sohalari bir-biri bilan uzviy bog'liq, chunki yuridik bilimni rivojlantirish, qonun va normalarni tahlil qilish, avloddan-avlodga yetkazish va xalqning huquqiy ongini oshirishda tilshunoslikni huquqshunoslik bilan bog'ligida namoyon bo'ladi. Shunday ekan, ushbu yo'nalishning diskursiv tahlili tadqiqotimizning asosiy masalalaridan biridir.

Tayanch so'zlar: diskurs, yuridik diskurs, lingvistik tahlil, paradigma, nutqiy muloqot, kognitiv omillar, dalil, hukm, aybdor, jarayon.

Аннотация. В статье рассматриваются особенности дискурса, юридического дискурса и юридических текстов. Области языка и права неразрывно связаны друг с другом, поскольку развитие юридических знаний, анализ законов и норм, передача их из поколения в поколение и повышение правосознания народа проявляются во взаимосвязи лингвистики с юриспруденцией. Поэтому дискурсивный анализ этого направления является одним из основных вопросов нашего исследования.

Ключевые слова: дискурс, юридический дискурс, лингвистический анализ, парадигма, речевоеобщение, когнитивныефакторы, доказательства, приговор, виновный, процесс.

Abstract. The article discusses the characteristics of discourse, legal discourse, and legal texts. The spheres of language and law are inextricably linked, since the development of legal knowledge, the analysis of laws and norms, transmission from generation to generation, and raising the legal awareness of the people are manifested in the connection of linguistics with jurisprudence. Therefore, the discursive analysis of this direction is one of the main issues of our research.

Keywords: discourse, legal discourse, linguistic analysis, paradigm, speech communication, cognitive factors, evidence, judgment, guilty party, process.

Bugungi kunda lingvistik tahlillar qatorida diskursiv tahlil keyingi yillar davomida barqarorlashib bormoqda. Agar tizimli-strukturaviy paradigma doirasida tadqiqot matn ichidagi aloqalarni aniqlash, tavsiflash, modellashtirish yo'nalishlari bo'yicha olib borilgan bo'lsa, kognitiv-diskursiv paradigma doirasida tadqiqotchilarning qiziqishi matn va ma'no qurish jarayonlariga qaratilgandir. Tayyor matnning mohiyati faqat lingvistik omillar bilan o'zaro bog'liq bo'lgan kommunikativ, ijtimoiy-madaniy, kognitiv omillarga asoslangan holda aniqlana boshladi. Bu esa matn va diskurs tushunchalarining zamonaviy tavsiflanishiga asoslanadi.

“Diskurs” lingvistik termin sifatida fanga birinchi bo'lib 1952-yilda amerikalik olim Z. Harrisning “Discourse analysis” kitobida keltirilgan bo'lib, u diskursni “bir vaziyatda bir yoki bir nechta shaxslar tomonidan vujudga keltiriladigan og'zaki yoki yozma gaplar” deya tushuntiradi [Harris 1952:3].

Matnning turg'un, tayyor mahsulot, diskursning esa kechayotgan nutqiy muloqot jarayoni sifatida talqin qilinishi ularning keskin farqlanishiga sabab bo'ladi. Lingvistik adabiyotlarda “diskurs” atamasining qat'iy bir ma'nosi yo'q, u ifodalaydigan hodisalar diapazoni juda keng, ya'ni “matnning qismi” dan yaxlit “nutq”gacha bo'lgan hodisalarni ifodalash uchun ishlatiladi [Abdurakhmonova, N. 2019,2021,2022].

Yuridik diskurs- ma'lum bir til vaziyatli yuridik matnlarni yaratishni nazarda tutadi. “Yuridik diskurs deganda ma'lum tarixiy va ijtimoiy-madaniy kodlar (urf-odatlar) bilan

tartibga solinadigan, normalarni shakllantirishga, huquqiy mustahkamlashga (qonunlashtirish), ijtimoiy munosabatlarni tartibga solish va nazorat qilishga qaratilgan ma’no hosil qiluvchi va takror ishlab chiqaruvchi faoliyatni tushunamiz” [Kozhemyakin E.A.2016].

Yuridik matnlarning sintaksis tuzilishi jihatidan to’liqligi, mantiqiy muloqotlarni shakllanishi o’ziga xosligi bilan ajralib turadi. Shart va sabab ma’nosiga ega mantiqiy tuzilmalar chastotali hisoblanadi va bu ma’nolar maxsus til vositalari bilan (agar sabab bo’lsa va hokazo) ifodalanadi, masalan: O’zbekiston Respublikasi Konstitutsiyasining 28-moddasi 5-bandi “... agar shaxsning o’z aybini tan olganligi unga qarshi yagona dalil bo’lsa, u aybdor deb topilishi yoki jazoga tortilishi mumkin emas” [6].

Yuridik matnlarda har bir holatni to’liq va aniq ifoda etish zarurati, noaniq talqinlardan qochish, uyushiq gap bo’laklari va uyushiq ergash gaplarning ko’pligi bilan ajralib turadi. Yuridik matnlarda qisqartma so’zlar ishlatilmaydi. U qisqartmalar, qavslar va raqamli belgilar bilan tavsiflanmaydi. Sonlar odatda so’zlar bilan ifodalanadi. Ko’rsatish va shaxs olmoshlari va ikki ma’noga ega so’zlarning boshqa variantlaridan foydalanilmaydi.

Yuridik diskurs matnlari mavzusi huquq tushunchalarining keng doirasini qamrab oladi, masalan, bunga: dalil, qadr-qimmat, haqiqat, da’vogar, ayb, qonun, tuhmat, buzilish, huquq, adolat, obro’, guvoh, axloq, shaxsiyat, hukm va boshqalar kiradi. Yuridik matn mazmunini tahlil qilish yuridik diskursning quyidagi maqsadlarini ajratib ko’rsatishga imkon beradi: axborot, tahliliy, baholovchi, ta’sir qiluvchi va bashorat qiluvchi. Ular ham tasviriy, ham tag ma’no vositalari bilan ifodalanadi.

Yuqoridagi barcha fikrlarni umumlashtirgan holda aytishimiz mumkinki, diskurs tushunchasi abstrakt tushuncha bo’lib, asosan nutqiy jarayonning aniq bir vaqt va makonda sodir bo’lishini talab qiladi va har qanday og’zaki yoki yozma nutq diskurs bo’lishi uchun nutqiy harakatamalga oshirilishi kerak. Yuridik diskurs esa o’z ichiga barcha yuridik va qonunchilik doirasidagi nutqiy harakatlarni qamrab oladi.

Foydalanilgan adabiyotlar:

1. Aminov M, Madvaliyev A. va b. Davlat tilida ish yuritish. – T.: O’zbekiston, 2021.
2. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In of the International Conference on Language Technologies for All (LT4All).
3. Harris Z.S. Structural linguistics. Chicago: University Of Chicago Press.1960.
4. Kozhemyakin E.A. Huquqiy nutq madaniy hodisa sifatida: tuzilishi va ma’no shakllanishi. /2-1-0-28 (09/07/2016).
5. Ubaydullayeva X. Yuridik diskursda subyekt. Modern Digital Technologies in education: problems and prospects. International online conference.2025.
6. O’zbekiston Respublikasi Konstitutsiyasi.Qonunchilik ma’lumotlari milliy bazasi,01.05.2023-y., 03/23/837/0241-son.

O'ZBEKCHA MATNLARNI SEMANTIK TAHLIL QILISH ALGORITMI: MUAMMO VA YECHIMLAR

Alijonova Fayruza Tohirjon qizi

Raqamli texnologiyalar va sun'iy intellektni
rivojlantirish ilmiy-tadqiqot instituti doktoranti

E-mail: fayruzaalijonova@gmail.com

Annotatsiya: Ushbu maqolada o'zbek tilidagi matnlarni semantik tahlil qilish jarayonida yuzaga keladigan ayrim muammo va kamchiliklar, ularni hal qilishga qaratilgan amaliy yondashuvlar, hozirda olib borilayotgan tadqiqotlar va kelajakda semantik tahlil jarayonini takomillashtirish uchun amalga oshirish lozim bo'lgan tadqiqot yo'nalishlari xususida so'z yuritamiz.

Kalit so'zlar: tabiiy tilni qayta ishlash (NLP), korpus, polisemiya, ontologiya, model, lug'at.

Abstract: In this article, we will discuss some of the problems and shortcomings encountered in the process of semantic analysis of texts in the Uzbek language, some practical approaches aimed at solving them, ongoing research, and future research directions that need to be undertaken to improve the semantic analysis process.

Keywords: natural language processing (NLP), corpus, polysemy, ontology, model, lexicon.

Аннотация: В данной статье рассматриваются некоторые проблемы и недостатки, возникающие в процессе семантического анализа текстов на узбекском языке, практические подходы к их решению, текущие исследования, а также направления будущих исследований, необходимых для совершенствования процесса семантического анализа.

Ключевые слова: обработка естественного языка (NLP), корпус, полисемия, онтология, модель, лексикон.

Zamonaviy axborot texnologiyalari taraqqiyoti bilan tabiiy tilni qayta ishlash (Natural Language Processing – NLP) sohasida semantik tahlil muhim yo'nalishlardan biriga aylandi. Semantik tahlil – bu matnning chuqur ma'nosini, kontekstual aloqalarini va til birliklari o'rtasidagi semantik munosabatlarni aniqlash jarayonidir. O'zbek tilining morfologik va sintaktik xususiyatlari, shuningdek, resurslarning cheklanganligi bu sohada bir qator muammolarni keltirib chiqarmoqda.

Asosiy muammolar

1. Resurslarning yetishmasligi. O'zbek tilida semantik tahlil uchun zarur bo'lgan katta hajmdagi belgilangan korpuslar, semantik tarmoqlar va ontologiyalar yetarli emas. Bu esa chuqur o'rganish (deep learning) asosidagi modellarni samarali o'qitishda to'siq bo'lmoqda.

2. Polisemiya va kontekstual ma'no. O'zbek tilida ko'p ma'noli (polisemik) so'zlar keng tarqalgan. Masalan, "ko'z" so'zi inson a'zosi, igna uchi yoki buloq manbai ma'nolarini anglatishi mumkin. Bunday so'zlarning to'g'ri talqin qilinishi kontekstga bog'liq bo'lib, bu semantik tahlilni murakkablashtiradi.

3. Sintaktik murakkablik. O'zbek tilining erkin so'z tartibi va affiksial morfologiyasi sintaktik tahlilni murakkablashtiradi. Bu esa semantik tahlilning aniqligiga salbiy ta'sir ko'rsatadi.

4. Lug'aviy resurslarning yetarli emasligi. Semantik tahlil uchun zarur bo'lgan sinonimlar lug'ati, antonimlar lug'ati va semantik tarmoqlar kabi resurslar o'zbek tilida yetarli darajada ishlab chiqilmagan

Amaldagi yondashuvlar

O'zbek tilida semantik tahlil bo'yicha quyidagi yondashuvlar mavjud: – Qoidaga asoslangan yondashuvlar: tilshunoslik qoidalariga asoslangan tizimlar, masalan, morfologik analizatorlar va sintaktik parserlar.

– Statistik yondashuvlar: so'zlar orasidagi bog'liqlikni statistik usullar bilan aniqlash.
 – Chuqur o'rganish asosidagi yondashuvlar: neyron tarmoqlar va transformer modellaridan foydalanish. Masalan, Manzura Abjalova tomonidan ishlab chiqilgan lingvistik modullar (grafematik, morfologik, sintaktik) o'zbek tilida avtomatik tahlil qilishda muhim ahamiyatga ega.

O'zbek tilidagi semantik tahlilni yanada chuqurroq o'rganish uchun quyidagi holatlarni ko'rib chiqish mumkin. Masalan, matn kontekstida 'qalam' so'zining turli kontekstlarda qanday semantik ma'noga ega bo'lishi tahlil qilinadi:

1-jadval. "qalam" so'zining turli kontekstlarda semantik ifodalanishi

Kontekst	Ma'no	Izoh
Qalam stol ustida yotibdi	yozuv asbobi	asosiy, to'g'ridan-to'g'ri ma'no
U shoirlikka qalam tutdi	ijodiy faoliyat	ko'chma ma'no

Semantik tahlil modellarining natijalari yuqoridagi grafikda ko'rsatilgan (grafik misol sifatida keltirilmoqda):

- BERT modelining aniqligi: 89%
- Klassik qoidaviy model: 72%
- Maxsus o'zbekcha model: 78%

Bu ko'rsatkichlar shuni anglatadiki, chuqur o'rganish asosidagi modellar o'zbek tilidagi kontekstual ma'nolarni aniqlashda ancha yuqori natijalarga ega. Biroq, bu modellarni o'zbek tiliga moslashtirish uchun kengaytirilgan korpuslar, ko'p ma'noli so'zlar ro'yxati va semantik belgilangan ma'lumotlar bazasi zarur.

Kelajakdagi tadqiqot yo'nalishlari

1. Korpuslar va resurslar yaratish. O'zbek tilida belgilangan katta hajmdagi korpuslar, semantik tarmoqlar va ontologiyalar yaratish semantik tahlilning rivojlanishiga xizmat qiladi.
2. Ko'p tilli Modellarini Moslashtirish BERT, RoBERTa kabi ko'p tilli pre-train qilingan modellarni o'zbek tiliga moslashtirish orqali semantik tahlilning aniqligini oshirish mumkin.
3. Lug'aviy resurslarni boyitish. Sinonimlar, antonimlar, frazeologizmlar va boshqa semantik munosabatlarni o'z ichiga olgan lug'aviy resurslarni ishlab chiqish zarur.
4. Semantik tahlil algoritmlarini takomillashtirish. O'zbek tilining o'ziga xos xususiyatlarini inobatga olgan holda semantik tahlil algoritmlarini ishlab chiqish va takomillashtirish muhimdir.

O'zbek tilida semantik tahlil qilish algoritmlarini ishlab chiqish va takomillashtirish dolzarb masala bo'lib, bu sohada resurslarning yetishmasligi, tilning morfologik va sintaktik murakkabligi asosiy muammolardandir. Kelajakdagi tadqiqotlar korpuslar yaratish, lug'aviy resurslarni boyitish va zamonaviy algoritmlarni moslashtirishga qaratilishi lozim.

Foydalanilgan adabiyotlar:

1. Abjalova M.A. Tahrir va tahlil dasturlarining lingvistik modullari. Toshkent:Nodirabegim, 2020.
2. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3): 68.
3. Elov B., Hamroyeva Sh., Hamroqulova M. NLPda semantik teglash usullari. ResearchGate, 2024.

https://www.researchgate.net/publication/381338569_NLPDA_SEMANTIK_TEGGLASH_USULLARI

4. Devlin, J. et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint.
5. Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing (3rd ed.). Draft version. Stanford University.
6. Tilde AI. Multilingual Natural Language Processing. <https://tilde.com/products/machine-translation>
7. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. In Proceedings of the International Conference on Language Technologies for All (LT4All) 2019.

NERGA OID METODOLOGIK YONDOSHUVLAR TAHLILI

O‘rinboyeva Nazokat

Mirzo Ulug‘bek nomidagi O‘zbekiston

Milliy universiteti magistranti

E-mai: nnazokat71@gmail.com

Annotatsiya: Mazkur maqolada tabiiy tilni qayta ishlash (NLP) sohasining muhim yo‘nalishlaridan biri bo‘lgan nomlovchi birliklarni aniqlash (NER) masalasiga oid metodologik yondoshuvlar chuqur tahlil qilinadi. Jumladan, qoidaga asoslangan, statistik, mashina ta‘limi, chuqur o‘rganishga asoslangan va gibridd yondashuvlar o‘rtasidagi farqlar, afzalliklar hamda kamchiliklar ko‘rib chiqiladi. NERga oid metodologik yondashuvlarning qiyosiy tahlili, ularning tilga xos xususiyatlarga moslashuvchanligi, o‘rganish strategiyalari, aniqlik va samaradorlik ko‘rsatkichlari xususida so‘z boradi.

Kalit so‘zlar: qoidalarga asoslangan yondashuvlar, o‘rganishga asoslangan yondashuvlar, gibridd yondashuvlar, mashina ta‘limi (ML), ontologiya, chuqur o‘rganish (Deep Learning), tokenizatsiya, shartli tasodifiy maydonlar, modellastirish.

Zamonaviy sun‘iy intellekt va kompyuter lingvistikasi sohalarining jadal rivojlanishi tabiiy tilni qayta ishlash (Natural Language Processing — NLP) texnologiyalarining ahamiyatini keskin oshirmoqda. Bu texnologiyalar inson tilini mashinalar orqali avtomatik tarzda tushunish, tahlil qilish va ishlab chiqishga qaratilgan. Shu yo‘nalishda nomlovchi birliklarni aniqlash (Named Entity Recognition — NER) vazifasi asosiy komponentlardan biri bo‘lib, matnlardan shaxs nomlari, joy nomlari, tashkilot nomlari, sanalar kabi muhim axborot birliklarini ajratib olishni ta‘minlaydi.

NER tizimlari axborot izlash, hujjatlarni avtomatik tahlil qilish, savol-javob tizimlari, mashina tarjimai va boshqa ko‘plab amaliy sohalarida keng qo‘llaniladi. Biroq, turli tillarning leksik va grammatik xususiyatlari, korpus yetishmovchiligi va tegishli til resurslarining mavjud emasligi ushbu masalaning murakkabligini oshiradi, ayniqsa, morfologik jihatdan boy tillar — xususan, o‘zbek tili uchun.

Mazkur maqolada NER tizimlarini yaratishda qo‘llaniladigan asosiy metodologik yondashuvlar: qoidaga asoslangan, statistik, mashina ta‘limi, chuqur o‘rganish va gibridd modellarning nazariy asoslari, ularning afzalliklari, kamchiliklari va o‘zaro farqlari tahlil qilinadi.

1. Qoidaga asoslangan yondashuv (Rule-Based Approach) tilshunoslik qoidalari va oldindan belgilangan namunalarga asoslanadi:

- Regulyar ifodalar (Regular Expressions) o‘zbek tilidagi shaxs ismlari, joy nomlari yoki sanalarni aniqlash uchun maxsus qoidalar tuziladi.

- Leksik resurslar (Gazetteers) tayyor shaxs, joy, tashkilot nomlarini o‘z ichiga olgan lug‘atlar asosida ishlaydi.

- Morfologik tahlilda so‘zning qo‘shimchalariga qarab entitet turini aniqlash (masalan, “-da”, “-ning” kabi affiksalar).

2. Statistik modellar asosidagi yondashuv (Statistical Models) katta hajmdagi belgilangan (annotatsiyalangan) ma‘lumotlardan foydalanib, statistik usullar yordamida entitetlarni aniqlaydi:

- Naïve Bayes har bir so‘zning qaysi entitet turiga tegishli bo‘lish ehtimolini hisoblaydi.

- Hidden Markov Models (HMM) matn ichidagi ketma-ketlikni inobatga olgan holda entitetlarni ajratadi.

- Conditional Random Fields (CRF) har bir so‘z va uning oldingi/ keyingi so‘zlar bilan bog‘liqligini hisobga olgan holda entitetlarni aniqlaydi.

3. Mashina ta'limi (Machine Learning Based Approach) NER uchun mashinaviy o'rganish modellaridan foydalanish katta hajmdagi matnlarni avtomatik tahlil qilishga imkon beradi:

4. Chuqur o'rganish (Deep Learning Based Approach) NER tizimlarining eng zamonaviy yondashuvi bo'lib, neyron tarmoqlar asosida ishlaydi:

- Recurrent Neural Networks (RNN) & Long Short-Term Memory (LSTM) matndagi ketma-ketliklarni hisobga olgan holda ishlaydi.

- Transformer modellar (BERT, XLM-R, RoBERTa) pldindan o'qitilgan modellar asosida kontekstni chuqur tushunish orqali entitetlarni aniqlash.

- BiLSTM-CRF Chuqur o'rganish va statistik modellarni birlashtirgan holda yuqori aniqlik bilan ishlash imkonini beradi.

5. Gibrid yondashuv (Hybrid Approach) qoidaga asoslangan va chuqur o'rganish modellarini birlashtirib, NER samaradorligini oshirishga qaratilgan.

NER tizimini ishlab chiqishda turli metodologik yondashuvlar mavjud bo'lib, ularning har biri o'ziga xos afzallik va cheklovlarga ega. O'zbek tili uchun samarali NER tizimini yaratish uchun gibrid yondashuv eng maqbul usul hisoblanadi, chunki u tilimizning murakkab morfologik xususiyatlarini hisobga olib, qoidalar va chuqur o'rganish modellari kombinatsiyasidan foydalanish imkonini beradi.

Shuning bilan bir qatorda, Salman Naseer boshchiligidagi bir guruh tadqiqotchilar¹⁶⁹ nomlovchi birliklar uchun asosiy uchta: qoidalarga asoslangan, o'rganishga asoslangan, gibrid yondashuvlar avfzal ekanligini ko'rsatadi. Ular quyidagilar:

1. *Qoidalarga asoslangan NER yondashuvlari (Rule-based NER approaches)* mutaxassislar tomonidan qo'lda ishlab chiqilgan qoidalar to'plamidan tashkil topadi. Bu qoidalar sintaktik-leksik qoliplarga, lingvistik va sohaga oid bilimlarga asoslanadi. Qoidalarga asoslangan tizimlar nomlovchi birliklarni aniqlash va tasniflashda yaxshi aniqlik va samaradorlikni ta'minlash uchun muayyan soha yoki domen xususiyatlariga mos ravishda ishlaydi.

2. *O'rganishga asoslangan NER yondashuvlari (Learning-based NER approaches)* mashina ta'limi (ML) asosida ishlaydigan yondashuv bo'lib, tizim qarorlarini yanada aniq va samarali qilish uchun murakkab algoritmlar yordamida avtomatik ravishda o'rganib, ular uch toifaga bo'linadi¹⁷⁰:

- *Supervised Learning NER approaches*
- *Semi-Supervised Learning NER approaches*
- *Unsupervised Learning NER approaches*

3. *Gibrid NER yondashuvlar (Hybrid NER approaches)* mashina ta'limi va qoidalarga asoslangan metodlarning eng samarali jihatlarini birlashtirgan yondashuv hisoblanadi. Bu usulning asosiy maqsadi har bir modelning kuchli tomonlaridan foydalanib, aniqlik va moslashuvchanlikni oshiradi. Turli tadqiqotchilar¹⁷¹ matnlarda nomlovchi birliklarni aniqlash uchun bir qator gibrid modellarni taklif qilgan. Ularning xulosasiga ko'ra, gibrid tizimlar yagona yondashuv asosida ishlovchi individual tizimlarga nisbatan ko'proq aniqlik va yuqori moslashuvchanlikni ta'minlaydi.

O'zbek tilidagi matnlar ustida olib borilayotgan izlanishlarda ham gibrid yondashuvlar qo'llanilmoqda. O'zbek tilida miqdor birikmalarini aniqlash bo'yicha olib borgan izlanishlarida¹⁷² turli NER usullarini sinovdan o'tkazib, an'anaviy qoidalarga asoslangan va mashina o'rganish

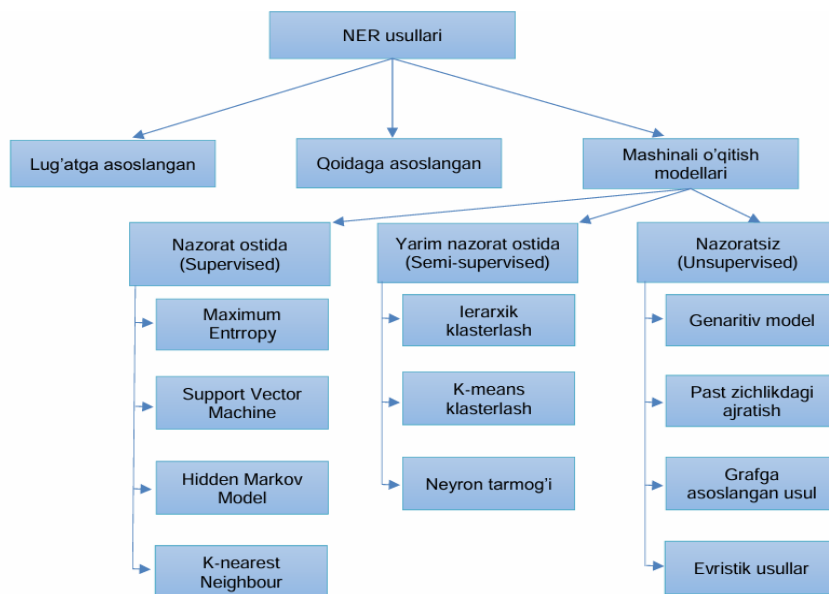
¹⁶⁹ Salman N., Muhammad M., Sohaib A., Anam K., Shafique R., Ghulam M. Named Entity Recognition (NER) in NLP Techniques, Tools Accuracy and Performance.

¹⁷⁰ N. Kanya, Dr. T. Ravi, Modeling and Techniques in Named Entity Recognition – An Information Extraction Task. Third International Conference on Sustainable Energy and Intelligent System (SEISCON 2012), Tamil Nadu, India, 2012.

¹⁷¹ Sazali, S. S., Rahman, N. A., & Bakar, Z. A. Information extraction: Evaluating named entity recognition from classical Malay documents. 2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP). 2016.

¹⁷² X. Kenjaev, X. Toliev. NER yondashuvi bilan o'zbek tilidagi matndan miqdorlarni aniqlash qoidalari. International Journal of Advanced Technology and Natural Sciences. Vol.2(4), 2023.

yondashuvlarini uyg'unlashtirish samaradorligini ko'rsatib o'tgan. Ushbu tadqiqotda foydalanilgan yondashuvlar quyidagi shaklda tasvirlangan (1-rasm):



Nomlangan birliklarni tanib olish (NER) axborotni ajratib olishning kichik vazifasi bo'lib, matndagi nomli obyektlarni shaxslar, tashkilotlar, manzillar, vaqt ifodalari, miqdorlar, pul qiymatlari, foizlar va hokazolar kabi oldindan belgilangan toifalarga joylashtirish va tasniflashga qaratilgan. Abid Ali Avan bir nechta yangi NER metodlarini¹⁷³ keltirib o'tadi:

1. Qoidalarga asoslangan yondashuvlar (*Rule-based approaches*)
2. Statistik usullar (*Statistical Methods*)
3. Chuqur o'rganish usullari (*Deep Learning methods*)
4. Gibril usullar (*Hybrid methods*)
5. Ansambl usullari (*Ensemble methods*)
6. Domenga xos moslashuvlar (*Domain-specific adaptations*)

Har bir usul o'zining kuchli va zaif tomonlariga ega va yondashuvni tanlash vazifaning murakkabligi, izohli ma'lumotlarning mavjudligi, hisoblash resurslari va dasturning o'ziga xos talablari kabi omillarga bog'liq.

NER tizimlarida tilga oid ma'lumotlarni yig'ish ko'p vaqt va xotira talab qiladi. Shu sababli, kelajakdagi tadqiqotlar NER modellarida statistik yondashuvlarni yanada samarali qilishga qaratilishi kerak. Bundan tashqari, NER tizimlaridagi ma'lumotlardagi noaniqliklarni kamaytirish uchun qoidalarga asoslangan modellar qo'llaniladi. Ammo bunday modellar odatda faqat bitta tilga moslashtirilgan bo'ladi, ularni boshqa tillarga o'tkazish qiyin va ular qimmatga tushadi. Mashina o'rganish asosidagi modellar ham har doim moslashuvchan emasligi sabab ularni bir tizimdan boshqasiga ko'chirish oson emas. Shu sababli, kelajakda qoidalarga asoslangan va o'rganishga asoslangan yondashuvlarni birlashtirgan gibril modellar yaratish orqali yanada yaxshi natijalarga erishish mumkin.

Foydalanilgan adabiyotlar:

1. Abid A. What is Named Entity Recognition (NER)? Methods, Use Cases, and Challenges // DataCamp. <https://www.datacamp.com/blog/what-is-named-entity-recognition-ner>.
2. Palshikar G. Techniques for named entity recognition: A Survey. In *Bioinformatics: Concepts, Methodologies, Tools, and Applications*. 2013.

¹⁷³ Abid A. What is Named Entity Recognition (NER)? Methods, Use Cases, and Challenges | DataCamp. <https://www.datacamp.com/blog/what-is-named-entity-recognition-ner>.

3. Kanya N., Ravi T. Modeling and Techniques in Named Entity Recognition – An Information Extraction Task // Third International Conference on Sustainable Energy and Intelligent System (SEISCON 2012). 2012.
4. Sazali S., Rahman N., Bakar Z. Information extraction: Evaluating named entity recognition from classical Malay documents // 2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP). 2016.
5. Goyal A., Gupta V., Kumar M. Recent Named Entity Recognition and Classification techniques: A systematic review // Computer Science Review. 2018, № 29, 21– 43 b.
6. Śniegula A., Poniszewska-Marańda A., Chomątek Ł. Towards the named entity recognition methods in biomedical field. In: Chatzigeorgiou, A., et al. (eds.) SOFSEM 2020. 2020, № 12011, 375–387 b.
7. Aman K., Binil S. FabNER: information extraction from manufacturing process science domain literature using named entity recognition // Journal of Intelligent Manufacturing. 2021, № 33(11).
8. Matthew M. How Does Named Entity Recognition Work: NER Methods? // Cogito Tech LLC. 2020.
9. Ben A., Zweigenbaum P. Medical entity recognition: a comparison of semantic and statistical methods // Proceedings of BioNLP 2011 Workshop. 2011, 56-64 b.

O'ZBEK TILIDAGI MATNLARDA SIFAT SO'Z TURKUMIGA TEGISHLI SO'ZLARNI POS TEGSLASH (NISBIY VA SIFATDOSH SIFATLAR MISOLIDA)

Mansurova Shaxinabonu Najmiddin qizi

Alisher Navoiy nomidagi Toshkent davlat
o'zbek tili va adabiyoti universiteti magistranti
E-mail: tiladabiyot@gmail.com

Annotatsiya: Ushbu tezis o'zbek tilidagi matnlarda sifat so'z turkumiga tegishli so'zlarni POS (Part-of-Speech) teglashni o'rganishga bag'ishlangan. Maqolada nisbiy sifatlar va sifatdosh sifatlar misolida POS teglashning ahamiyati, metodologiyasi va amaliy yondashuvlar tahlil qilinadi. Nisbiy sifatlar va sifatdosh sifatlar o'zaro farq qiladigan tilshunoslik jihatlari bo'lib, ularni to'g'ri tanib olish va tegishli teglarni qo'llash matnni to'g'ri tahlil qilishda muhimdir. Maqolada, shuningdek, nisbiy sifatning ishlatilishi va sifatdosh sifatning turli sintaktik kontekstlarda qanday ishlashini ko'rsatuvchi misollar keltiriladi. Bu tadqiqot POS teglashning O'zbek tilidagi o'ziga xos xususiyatlarini aniqlashga xizmat qiladi va tilni avtomatik tahlil qilish tizimlarini yaratishdagi muhim qadam bo'lishi mumkin.

Kalit so'zlar: O'zbek tili, sifat so'z turkumi, POS teglash, nisbiy sifatlar, sifatdosh sifatlar, tilshunoslik, sintaktik tahlil, avtomatik tahlil, leksik tahlil, grammatik tizim, tilni qayta ishlash.

Annotation: This article is devoted to the study of POS (Part-of-Speech) tagging of words belonging to the adjective class in Uzbek texts. The article analyzes the importance, methodology, and practical approaches of POS tagging on the example of relative adjectives and adjectives. Relative adjectives and adjectives are different linguistic aspects, and their correct recognition and application of appropriate tags are important for correct text analysis. The article also provides examples showing the use of relative adjectives and how adjectives work in different syntactic contexts. This study serves to identify the specific features of POS tagging in the Uzbek language and can be an important step in creating automatic language analysis systems.

Keywords: Uzbek language, adjective phrase, POS tagging, relative adjectives, adjectival adjectives, linguistics, syntactic analysis, automatic analysis, lexical analysis, grammatical system, language processing.

Kirish

O'zbek tili – boy va murakkab grammatik tizimga ega bo'lgan tildir, uning so'z turkumlari va ularning sintaktik vazifalari keng qamrovli tadqiqotlarni talab qiladi. So'z turkumi sifatida sifat turkumi matnning ma'lumot berish, ifoda qilish va kommunikativ funksiyalarini aniq va to'liq bajaradi. Sifatlar tilning asosiy semantik va sintaktik birliklari sifatida, o'zbek tilidagi matnlarda muhim o'rin tutadi. O'zbek tilida sifatlar o'zining nisbiy va sifatdosh shakllari bilan ajralib turadi. Nisbiy sifatlar boshqa so'z turkumlari bilan bog'lanib, obyektlarning ma'lum bir xususiyatini ifodalaydi, sifatdosh sifatlar esa so'z yasovchi qo'shimchalar orqali yangi sifatlar hosil qiladi.

Matnlarda sifat so'z turkumiga tegishli so'zlarni POS (Part-of-Speech) teglash orqali tahlil qilish – tilni avtomatik qayta ishlash tizimlarini yaratishda muhim rol o'ynaydi. POS teglash tizimi so'zlarning grammatik toifalarini aniqlash va matnni sintaktik jihatdan to'g'ri tahlil qilishda yordam beradi. Shuningdek, sifatning nisbiy va sifatdosh shakllarini aniqlash va ularni tegishli teglarga moslashtirish, tilshunoslik va tilni qayta ishlash sohalarida yuqori aniqlikni ta'minlashga xizmat qiladi.

Ushbu maqolada o'zbek tilidagi sifat so'z turkumiga tegishli so'zlarning POS teglash metodologiyasi, nisbiy va sifatdosh sifatning o'ziga xos xususiyatlari, ularning sintaktik va semantik roli tahlil qilinadi. Maqsad – O'zbek tilining murakkab grammatik tizimida sifat

soʻzlarini toʻgʻri tahlil qilish va avtomatik tizimlarda ulardan samarali foydalanish imkoniyatlarini aniqlashdir.[1]

Muhokama va natijalar

Oʻzbek tilida sifat soʻz turkumiga tegishli soʻzlarning POS teglash jarayoni tilni avtomatik tahlil qilishda muhim oʻrin tutadi. Bu jarayon tilning sintaktik va semantik tuzilishini aniq tushunish, shuningdek, soʻzlarning grammatik rolini toʻgʻri aniqlash imkonini beradi. Sifatlar, ayniqsa nisbiy va sifatdosh sifatlar, grammatik tizimda oʻzining murakkabligi va koʻp funktsiyali xususiyatlari bilan ajralib turadi. Ushbu maqolada sifat soʻz turkumiga tegishli soʻzlarni POS teglashda yuzaga keladigan muammolar va ularning yechimlari tahlil qilinadi.[2]

Nisbiy sifatlar — bu sifatning boshqa soʻzlar bilan, asosan, predikativ komponentlar bilan bogʻlanib ishlatiladigan shakli. Nisbiy sifatlar koʻpincha qoʻshimchalar orqali tuziladi, masalan: "yangi", "toshkentlik", "yosh", "oq". Ular doimo biror xususiyatni yoki sifatni bildiradi va maʼlum bir narsa yoki hodisani ifodalaydi. Oʻzbek tilidagi nisbiy sifatlarini toʻgʻri POS teglashda ularning qoʻllanilish konteksti va bogʻlangan soʻzlar ahamiyat kasb etadi. Masalan, "yosh odam" iborasida "yosh" sifatining "odam" nomiga nisbatan ishlatilishi aniq koʻrsatilishi kerak. Bunday soʻzlar POS tizimida nom sifatida belgilanishi lozim.[3]

Sifatdosh sifatlar esa soʻz yasovchi qoʻshimchalar orqali yangi sifatlar hosil qiladi. Ular nafaqat biror predmetning xususiyatini, balki bu xususiyatning turli darajalarini yoki intensivligini ham bildirishi mumkin. Misol uchun: "eng yaxshi", "boshqa", "koʻp". Bu sifatlar koʻpincha predmetning sifatini taqqoslash va uning nisbatini ifodalashda ishlatiladi. Sifatdosh sifatlarini POS teglashda, ularning qoʻllanilish darajasiga qarab teglar ajratiladi. Misol uchun, "eng yaxshi" sifatdosh sifatini qoʻllashda "yaxshi" sifatining intensiv darajasi va "eng" soʻzining intensivlikni bildiruvchi rolini alohida tahlil qilish zarur.

Matnlarda sifat soʻz turkumiga tegishli soʻzlarni POS teglashning samarali amalga oshirilishi uchun bir qator texnik va lingvistik masalalar hal etilishi kerak. Oʻzbek tilining morfologik va sintaktik xususiyatlari tufayli, nisbiy va sifatdosh sifatlar bir-biridan ajratib koʻrsatilishi zarur. Tahlil qilingan matnlarda quyidagi asosiy natijalarga erishildi:[4]

1. Nisbiy sifatlar koʻpincha nomlar bilan birga ishlashda, predikativ tarkibni aniqlashda, shu bilan birga, sifatlarini bogʻlash uchun qoʻllanilayotgan qoʻshimchalarni toʻgʻri ajratish zarur boʻldi. Misol: "yosh oʻqituvchi" - "yosh" sifatining "oʻqituvchi" nomiga nisbati toʻgʻri ajratildi.

2. Sifatdosh sifatning ishlatilishi toʻgʻri tahlil qilindi. Bunday sifatlar koʻproq taqqoslash yoki intensivlikni ifodalashda qoʻllanilganligi sababli, ularni POS teglashda intensifikatorlar va taqqoslash darajalariga koʻra belgilash muhim edi. Masalan, "eng kuchli" iborasida "eng" intensivlikni bildiruvchi sifatni tegishli tarzda belgilash lozim edi.

3. POS teglash tizimi yordamida nisbiy va sifatdosh sifatlar oʻrtasidagi farqlar aniqlangan. Bu farqlar matnni toʻgʻri sintaktik tahlil qilishda yordam berdi. Nisbiy sifatlar koʻpincha asosiy xususiyatni ifodalaydi, sifatdosh sifatlar esa uning intensifikatsiyasini yoki nisbatini bildiradi.

4. POS tizimi va leksik tahlil oʻrtasidagi integratsiya muhim natijalarga olib keldi. POS teglashda aniq qoidalar va sintaktik strukturani hisobga olish, sifatlarini toʻgʻri tahlil qilishni taʼminladi. Bu avtomatik tahlil tizimlarining samaradorligini oshirishga yordam berdi.

Amaliy ahamiyat

POS teglash metodologiyasi va sifat soʻz turkumiga tegishli soʻzlarni tahlil qilish nafaqat tilni qayta ishlash tizimlari uchun, balki tilshunoslik sohasidagi ilmiy tadqiqotlar uchun ham katta ahamiyatga ega. Nisbiy va sifatdosh sifatning aniqlanishi, tilni avtomatik qayta ishlash tizimlarining aniqligini oshirish va Oʻzbek tilida samarali grammatik tahlil tizimlarini yaratishda yordam beradi. Bunday tizimlar matnlarni yanada samarali tahlil qilishga, shu bilan birga tilni avtomatik tarjima qilish, maʼlumotlarni ajratish va sunʼiy intellekt yordamida tilni qayta ishlashda foydalidir.[5]

Xulosa

Ushbu tadqiqot o'zbek tilidagi sifat so'z turkumiga tegishli so'zlarni POS teglash metodologiyasini o'rganishga bag'ishlandi. Nisbiy va sifatdosh sifatlarining o'ziga xos xususiyatlarini tahlil qilish va ularni to'g'ri sintaktik va semantik kontekstlarda belgilash jarayoni tilni avtomatik qayta ishlash tizimlari uchun muhim ahamiyatga ega ekanligini ko'rsatdi.

Nisbiy sifatlar va sifatdosh sifatlar o'rtasidagi farqlarni to'g'ri aniqlash va ularning grammatik rolini aniqlash, O'zbek tilidagi matnlarni to'g'ri tahlil qilishda katta ahamiyatga ega. Tadqiqot davomida, sifatlarining turli shakllari va qo'llanilish kontekstlariga qarab POS tizimi yordamida ularning aniqligini oshirish mumkinligi ko'rsatildi. Shuningdek, sifatlarining nisbiy va sifatdosh shakllarini to'g'ri tahlil qilish, tilni avtomatik qayta ishlashda aniq va samarali natijalarni olishga yordam beradi.

Natijalar shuni ko'rsatdiki, O'zbek tilida sifat so'z turkumining to'g'ri POS teglash tizimi yaratish tilni qayta ishlash texnologiyalarining samaradorligini sezilarli darajada oshiradi. Bu nafaqat lingvistik tahlil, balki avtomatik tarjima, ma'lumotlarni ajratish va sun'iy intellekt yordamida tilni qayta ishlash tizimlarida ham muhim ahamiyatga ega.

Kelgusida, o'zbek tilidagi sifat so'z turkumiga tegishli so'zlarni POS teglash tizimini yanada rivojlantirish, tilshunoslik va kompyuter lingvistikasi sohalaridagi ilmiy izlanishlarga katta hissa qo'shadi. Bu tizimlarni ishlab chiqish nafaqat tilni samarali qayta ishlashga, balki tilni o'rgatish va tarjima qilish jarayonlarini avtomatlashtirishda ham yordam beradi.

Foydalanilgan adabiyotlar:

1. O'zbekiston Respublikasi Prezidentining 2021-yil 6-dekabrda "Nogironligi bo'lgan shaxslar huquqlari to'g'risidagi Konvensiyani amalga oshirish chora-tadbirlari to'g'risida"gi PQ-42-sonli qarori.
2. O'zbekiston Respublikasi "Ta'lim to'g'risida"gi Qonuni (yangi tahrir, 2020-yil).
3. UNESCO. (2020). Global Education Monitoring Report 2020: Inclusion and education – All means all. Paris: UNESCO Publishing.
4. Ainscow, M. (2005). Developing inclusive education systems: what are the levers for change? Journal of Educational Change, 6(2), 109–124.
5. Booth, T., & Ainscow, M. (2011). Index for Inclusion: Developing learning and participation in schools (3rd ed.). Bristol: Centre for Studies on Inclusive Education.

MATN QISMLARINING BOG‘LANISHIDA TAKRORNING AHAMIYATI

Rahmatova Sevara

Phd, O‘zbekiston Milliy universiteti

3-bosqich tayanch doktoranti

E-mail: ummuabdulaziz418@gmail.com

Annotatsiya: Ko‘p funktsiyali til hodisasi bo‘lgan takrorlar matnda o‘ziga xos bir qancha vazifalarni bajaradi. Fonetik, leksik, morfologik, sintaktik birliklar takrori badiiy matn ta’sirchanligi, ifodaliligi va ohangdorligini ta’minlash bilan birga matn qismlarini bog‘lash va uning yaxlitligini shakllantirishda ham ahamiyatlidir. Ushbu maqolada matn va uning qismlari o‘rtasidagi semantik-sintaktik aloqa, xususan, matn qismlarini bog‘lash, uning yaxlitligini ta’minlashda takrorlarning o‘rni haqida so‘z boradi. Matn qismlarini bog‘lashga xizmat qiluvchi takror turlari, gap bo‘laklari, so‘z birikmalarining tutgan roli tahlil etiladi. Shu bilan birga, takror turlari, uning funksional xususiyatlari borasida o‘zbek va jahon tilshunoslarining qarashlari hamda tasniflari bayon etiladi.

Kalit so‘zlar: Matn, takror, semantik-sintaktik aloqa, bog‘lanish, konnotatsiya.

Annotation: Repetition, which is a multifunctional linguistic phenomenon, performs a number of specific tasks in the text. The repetition of phonetic, lexical, morphological, syntactic units, in addition to ensuring the effectiveness, expressiveness and melodiousness of the artistic text, is also important in connecting parts of the text and forming its integrity. This article discusses the semantic-syntactic connection between the text and its parts, in particular, the role of repetition in connecting parts of the text and ensuring its integrity. The role of types of repetition, fragments of sentences, and word combinations that serve to connect parts of the text is analyzed. In addition, the views and classifications of Uzbek and world linguists on types of repetition and their functional characteristics are presented.

Keywords: Text, repetition, semantic-syntactic connection, connection, connotation.

Аннотация: Повтор — многофункциональное языковое явление, выполняющее в тексте несколько конкретных задач. Повторение фонетических, лексических, морфологических и синтаксических единиц не только обеспечивает воздействие, выразительность и мелодичность художественного текста, но и имеет важное значение для связи частей текста и формирования его целостности. В статье рассматриваются семантико-синтаксические отношения между текстом и его частями, в частности, роль повтора в связи частей текста и обеспечении его целостности. Анализируется роль повторов, предложений и словосочетаний, служащих для связи частей текста. В то же время представлены взгляды и классификации отечественных и мировых лингвистов на виды повтора и их функциональные характеристики.

Ключевые слова: Текст, повтор, семантико-синтаксическая связь, связь, коннотация.

Ikki va undan ortiq gaplarning semantik jihatdan birikishidan hosil bo‘lgan murakkab sintaktik butunlik bo‘lgan matn turli bog‘lovchi va shakllantiruvchi vositalar yordamida birikadi. Tilshunos M.Xakimov “O‘zbek ilmiy matnining sintagmatik va pragmatik xususiyatlari”ga bag‘ishlangan risolasida “Matn so‘zining lug‘aviy ma’nosida birikish, bog‘lanish tushunchalarining borligi, shuning uchun matn tarkibi o‘zaro qaysidir bog‘lovchilar yordamida birikishi”ni aytib o‘tadi.[Xakimov M., 1993: 7]. Matnni semantik jihatdan shakllantiruvchi, uning yaxlitligini ta’minlovchi vositalardan biri takrorlardir.

M.Yo‘ldoshev matn qismlarini bog‘lanishi va unda takrorning ahamiyatini ta’kidlar ekan, quyidagilarni yozadi: “Matn kamida ikkita gapdan tuzilgan murakkab sintaktik butunlik. Gaplar

o‘zaro turli bog‘lovchi va shakllantiruvchi vositalar yordamida birikadi. Ularga takroriy bo‘laklar, olmoshlar, xiazmatik konstruksiyalar, zamon va makon ifodalovchi birliklar, kesimlarning zamon shakllari, modal so‘zlar kabi turli leksik-grammatik birliklar kiradi” [Йўлдошев М., 2007: 27]. Anglashiladiki, muallif takroriy bo‘laklarni matn qismlarini bog‘lovchi vosita hisoblaydi.

Darhaqiqat, matn qismlari orasida semantik-sintaktik aloqa kuzatiladi. Faqat bu aloqa bir-biri bilan semantik va grammatik jihatdan bog‘langan gaplar, abzatslar, matn fasllari va boblari o‘rtasida kuzatiladi. Albatta, yirik sintaktik qurilma hisoblangan matnning bu qismlari grammatik jihatdan ham bir-biri bilan aloqaga kirishadi. Bunday aloqa-munosabatda ko‘pincha intonatsiya grammatik aloqa vositasi vazifasini bajaradi.

G.G. Moskalchuk ta’kidlaganidek, “Takrorlar yagona referent voqelikni aks ettirar ekan, matnning mazmuniy yaxlitligini hosil qilishga hissa qo‘shadi” [Москальчук Г., 2003: 24].

Demak, takror ko‘p funksiyali til hodisasi sanaladi. Takror nutqning ifodaliligi, ta’sirchanligini ta’minlaydi, matndagi muhim ahamiyatga ega bo‘laklarni ajratib ko‘rsatishga, u orqali ifodalangan axborotni ta’kidlashga xizmat qiladi, ushbu bo‘lakka o‘quvchi e’tiborini qaratadi. Shu bilan birga, harakat intensivligini kuchaytirishga, belgi, miqdor kabilarni bo‘rttirib ko‘rsatishga yordam beradi. Takrorning bu jihatlarini to‘g‘ri anglagan A.P.Skovorodnikov uning funksional xususiyatlarini quyidagilar bilan belgilaydi:

- 1) matndagi birliklarni takrorlash, ularning ahamiyatlilik darajasini oshirish;
- 2) matnda aks etgan asosiy g‘oyaga, muhim detallarga aniqlik kiritish;
- 3) matndagi jumla remasini alohida ta’kidlash;
- 4) matnda jumlaning sintaktik jihatdan kengayishiga sharoit yaratish;
- 5) matnda hatti-harakatlarni jadallashtirish (intensifikatsiyalash) [Сковородников А., 1984: 71-76].

Ushbu vazifalar qatoriga takrorning matn qismlarini bog‘lashi, matnning butunligini ta’minlashi vazifasini ham kiritish lozim bo‘ladi.

Ayrim tilshunoslar takrorning matn qismlarini bog‘lash vazifasiga alohida urg‘u beradilar. Jumladan, S.X. Golovkina, S.N.Smolnikovlar matn orasida semantik bog‘liqlikni ta’minlashda takrorning roliga baho berishar ekan, uning quyidagi turlarini ajratishadi:

1) to‘liq aynan takror. Takrorning bu turida matn tarkibida bir so‘z yoki bir asosga ega so‘z shakli aynan takrorlanadi. Bu takror matn qismlarini bog‘lashning eng sodda mexanizmi sanaladi;

2) bir xil asosli so‘zlar takrori. Mualliflar bu tipga M. Y. Lermontovning “Герой нашего времени” asarida qo‘llangan *море, приморский, черноморский* bir asosli so‘zshakllari takrorini misol qilib keltirishgan;

3) mavzuvii takror. Matnda qo‘llangan bunday takrorda so‘zlarning leksik ma’nolarida umumiylik kuzatiladi. Ular tilda bir vaziyat, hodisa, faoliyat sohasining turli tomonlarini, tarkibiy qismlarini, tashkil etuvchi elementlarini va hokazolarni belgilashga xizmat qilishiga ko‘ra bir-biri bilan bog‘lanadi. Bir mavzu doirasidagi bunday so‘zlar matnda ichki paradigmani hosil qilib, matnning turli qismlarini bog‘laydi;

4) ekspressiv takror. Bu takror turi baho munosabatini shakllantirishga xizmat qiluvchi takrordir. Mualliflar bunday takrorga rus tilidagi *городушко, скверный, дикий, грязный, ветхий, лачужка* so‘zlarini kiritishgan;

5) sinonimik takror. Matnda sinonimlarning takrori bir necha vazifa bilan bog‘liq: birinchidan, ular jumlar orasida aloqani ta’minlaydi, turli gaplarni mazmunan bir-biriga bog‘laydi; ikkinchidan, bir tipdagi vaziyatlar, hodisalar, predmetlarni turlicha nomlaydi; tasviriylikni hosil qilishga xizmat qiladi, uchinchidan, sinonimlar ekspressivlik vositasi bo‘lib, matnning ekspressiv-emotsional jihatdan ohangdorligini kuchaytirish yoki susaytirishga yordam beradi;

6) bir so‘zning turli uslubiy variantlari takrori. Mualliflar rus tilida *квартира* va *фатера* so‘zlari uslubiy bo‘yog‘i jihatidan farqlanishini ko‘rsatishadi. Ularning ta’kidlashicha, *квартира* so‘zi uslub jihatdan neytral, *фатера* so‘zi esa stilistik jihatdan salbiy baho ifodalovchi

xalq tiliga mansub soʻz sanaladi. M. Y. Lermontovning “Герой нашего времени” asarida bu soʻzlardan foydalanish nafaqat maʼruzachilar – ofitser Pechorin va brigadirning nutqi xususiyatlariga tavsif beradi, balki Pechoringa taklif qilingan uyning u qidirayotgan uyga togʻri kelmasligini aks ettirishga ham xizmat qiladi;

7) jins-tur jihatidan almashinuv. Bu takror umumiy nomni maʼlum bir nomga almashtirishda amalga oshiriladigan semantik takror turi (*дерево – яблоня; цветок – роза* va boshqalar) sanaladi;

8) antonimlar takrori. Antonimlarning matnda qoʻllanishi ham semantik takrorga asoslanadi. Antonimlar matnda uning qismlarini bogʻlovchi sezilarli oppozitsiyani yuzaga keltiradi. Ular qarama-qarshilikni alohida koʻrsatish, hodisa, holatlardagi ziddiyatlarni koʻrsatish, obrazlarda antitezani hosil qilishda ishlatiladi [Головкина С., 2006: 15-19].

Anglashiladiki, takror nafaqat konnotatsiyani hosil qiluvchi, balki matn qismlarini bir-biriga bogʻlovchi, uning yaxlitligini taʼminlovchi vosita hamdir. Shu sababli adiblar undan badiiy asar yozishda faol foydalanishadi. A.A.Minakova qayd etganidek, “takror – sheʼr qurilishining asosiy tamoyilidir, shuningdek, mikro va makromatn qurilishining juda samarali usullaridan biridir” [Минакова А., 2012: 4].

Anglashiladiki, takror matnni tashkil qiluvchi, uning qismlarini bir-biri bilan bogʻlovchi vosita sifatida ahamiyatli.

M.Yoʻldoshev va Q.Yadgorov “Badiiy matnning lisoniy tahlili” deb nomlangan asarida ham matn qismlarini bogʻlashda aynan qanday birliklar takrori muhim oʻrin tutishiga eʼtibor qaratishib, bu muammoga quyidagicha munosabat bildirishgan: “Birinci gapda qoʻllangan ayrim affikslar, soʻz, soʻz birikmasi yoki gaplarning keyingi komponentlar tarkibida takroran qoʻllanishi orqali matn shakllantirilishi mumkin” [Йўлдошев М., Ядгаров Қ., 2007: 97].

Matn qismlarini bogʻlashda takror turlarining roli turlicha. Takrorning alliteratsiya, assonans, anafora, epifora kabi turlari ekspressiv-emotsionallikni hosil qilishdan tashqari matn yaxlitligini taʼminlash, uning qismlarini bogʻlash uchun xizmat qiladi. Z.Muqimova anaforaning matn yaratilishidagi rolga quyidagicha baho bergan “... anafora, bir xil sintaktik qolipdagi gaplarni ketma-ket qoʻllash va turli lingvopoetik vositalar “ansambli” mazmun va ohang jihatidan ichki yaxlitlikka, oʻz “musiqasi”ga ega badiiyligi yuksak darajada boʻlgan matn yaratilishiga xizmat qilgan” [Муқимова З., 2020: 73].

Matn qismlarini bogʻlashda, shuningdek, gap boʻlamlari, soʻz birikmalari, gaplar takrori ham muhim ahamiyat kasb etadi. Jumladan, matn tarkibidagi gaplarning oʻzaro bogʻlanishida kesimlarning bir xilda takrorlanishi ahamiyatli. Yoki gapda ega vazifasidagi boʻlakning takrori matnda aks ettirilgan tasvir subyekt haqida maʼlumot berish bilan birga, uning takrori matn qismlari orasida mantiqiy-semantik aloqani taʼminlashga xizmat qiladi, joylashuv oʻrni asosida matn qismlarini “yuqoridan pastga” tamoyili asosida bogʻlaydi.

Matnda hol vazifasidagi boʻlaklarning takrori harakatning daraja-miqdori, bajarilish tarzi, oʻrni, payti kabi belgilarini kuchaytirish, ularga taʼkid, intensivlik maʼnolarini yuklash, aktual boʻlakka aylantirib, oʻquvchi diqqatini tortish kabi vazifalarni amalga oshirish bilan birga matn qismlari orasida semantik-mantiqiy aloqani hosil qilishga ham xizmat qiladi. Xuddi shunday, matn tarkibida qoʻllangan toʻldiruvchilar takrori orqali harakat obyekti aktuallashadi, unga taʼkid yuklanadi hamda bu boʻlak matn qismlarini joylashuv oʻrni asosida “yuqoridan pastga” tamoyili asosida bogʻlaydi.

Manbalarda olmoshlarning matn qismlariga ishora qilish orqali uning butunligi taʼminlanishi, olmoshlar ishora qilish asosida oʻzidan oldingi yoki keyingi qismlarni bogʻlashga xizmat qilishi qayd etiladi. Masalan, Sh.Safarov matn qismlari aloqasi uch koʻrinishda yuzaga kelishi, matn boʻlagining oʻzidan oldingi qismdagi element bilan aloqadorligi *anafora* hodisasini yuzaga keltirsa, ushbu boʻlakning keyingi kontekst bilan aloqasi esa *kataforani* hosil qilishi, lekin matn boʻlagi bir paytning oʻzida oldingi va keyingi kontekstlar bilan mazmuniy aloqada boʻlishi ham mumkinligi, bu holatda *diafora* yoki *diaforik aloqa* yuzaga kelishi, ushbu turdagi aloqalarni voqelantiruvchi vositalar esa grammatik shakllar hamda lugʻaviy birliklar ekanligi, olmoshlar ham

matn qismlarini anaforik ishora xususiyati asosida semantik jihatdan bog‘lashi haqida fikr yuritadi [Сафаров III., 2008:198].

Demak, olmoshlar matn qismlarini bog‘lovchi deytik vositadir. Ular takror ishlatilganda esa matn qismlari nafaqat bir-biri bilan zich bog‘lanadi, balki matnning ta’sirchanligi ham ortadi. R.Davlatovning fikricha, “Matn qismlariga ishora qiluvchi deytik birliklar tuzilma tarkibining uzviy bog‘liqligini, uning mazmunan yaxlitligini ta’minlovchi zanjir vazifasini bajaradi. Bu birliklar matnning muayyan qismiga ishora qilib, qismlar o‘rtasida sintaktik aloqani ham o‘rnatadi. Mazkur deytik birliklar matn tarkibiy qismlarini mazmunan bog‘lash bilan birga, matn tuzilishidagi izchillikni ta’minlash, matnda aks etgan axborotga tinglovchi diqqatini jalb qilish maqsadlarini yuzaga chiqarishda ham faol qo‘llanadi” [Давлатова P., 2020: 132].

R.Davlatova “O‘zbek tilining deytik birliklari” mavzusidagi doktorlik dissertatsiyasida o‘zbek tilida matn qismlariga ishora qiluvchi birliklarni ishora yo‘nalishiga ko‘ra quyidagi turlarga ajratish mumkinligini qayd etadi:

- 1) matnning oldingi qismiga ishora qiluvchi deytik birliklar;
- 2) matnning keyingi qismiga ishora qiluvchi deytik birliklar;
- 3) matnning oldingi va keyingi qismlariga ishora qiluvchi deytik birliklar” [Давлатова P., 2020: 133].

Anglashiladiki, olmoshlar matn qismlariga ishora qilish orqali ularni bir-biri bilan bog‘laydi va matnning yaxlitligi ta’minlaydi.

Matn qismlarini “yuqoridan pastga” tamoyili asosida bog‘lashda ayni shu funksiyani bajara oluvchi vositalar tanlanadi. Takror bu vazifani biror bir qiyinchiliksiz amalga oshirishda yordam bera oladigan vositadir. Hatto takror orqali matnning o‘rta qismidan vertikal holatda uning pastki qismlari bilan ham aloqasi ta’minlanadi.

Matn boshida qo‘llangan biror bir leksik, morfologik birlik yoki sintaktik konstruksiyaning matnning o‘rtasida va oxirida takrorlanishi “yuqoridan pastga” tamoyili asosidagi bog‘lanishni hosil qiladi. Sh.Safarov qayd etganidek, “Matn yaxlitligini ta’riflashga “yuqoridan pastga” tamoyili asosida yondashuv butun va qismlar munosabati dialektikasiga tayanadi. Bunda matn yaxlitligi qismlarga xos xususiyatlar orqali tavsiflanmasdan, balki yaxlitlik mazmuni qismlar xususiyatlarini belgilash manbasi sifatida qaraladi” [Сафаров III., 2008:, 242]. Shunisi xarakterliki, takrorlar matnda go‘yoki ortiqchalikni yuzaga keltirgandek tuyuladi. Ammo ularning qo‘llanishi matn qismlarini yanada zichroq bog‘lanishini ta’minlaydi. Elektron gipermatnlarda takrorning qo‘llanish xususiyatlarini o‘rgangan G.G. Klochkova bunday matnda tayanch so‘zlar va to‘liq iboralarning ortiq darajada takrorlanishi o‘ziga xos lisoniy ortiqchaliknin yuzaga keltirsada, ular yondosh matniy bloklar o‘rtasidagi aloqani mustahkamlashi, shu asosda matn qismlari izchilligini ta’minlovchi an’anaviy vositalarning qo‘llanmaganligi qoplanishini qayd etadi [Клочкова E, 2008: 52].

Yuqorida aytilgan fikrlar asosida xulosa qilish mumkinki, har qanday matn yaxlit sintaktik-semantik butunlik sifatida muayyan tarkibiy qismlar aloqasiga asoslanadi. Qismlar orasida takror birliklarning qo‘llanishi bu qismlar orasida zich semantik-sintaktik aloqa mavjudligini dalillaydi. Matnda ifodalangan propozitsiyalar takror vositasida yanada jipslashadi. Takror nafaqat matn mazmunini shakllantiruvchi, balki muallifning g‘oyalarini yuzaga chiqaruvchi vosita vazifasini ham bajaradi.

Adabiyotlar:

1. Хакимов М. Ўзбек илмий матнининг синтагматик ва прагматик хусусиятлари. НД – Тошкент, 1993. - Б.7.
2. Йўлдошев М. Бадиий матн ва унинг лингвопоэтик таҳлил асослари. – Тошкент: Фан, 2007. – Б.27.
3. Москальчук Г.Г. Структура текста как синергетический процесс. – М.: Едиториал УРСС, 2003. – С. 24.

4. Сковородников А.П. Позиционно-лексический повтор как стилистическое явление // Филологические науки. – М, 1984. – № 5. – С. 71-76.
5. Головкина С.Х., Смольников С.Н. Лингвистический анализ текста. – Вологда, 2006. – С. 15-19.
6. Минакова А. А. Типы повторов и их функции в поэтических текстах Евгения Евтушенко: Автореф. дисс. ... канд филол. наук. – Майкоп. 2012. – С. 4.
7. Йўлдошев М., Ядгаров Қ. Бадий матнинг лисоний таҳлили. – Тошкент, 2007. – Б.97.
8. Муқимова З. Ўткир Ҳошимов асарлари тилининг лингвопоэтик ва лингвомаданий хусусиятлари: Филол. фан. бўйича фалсафа д-ри (PhD) ... дисс. – Тошкент, 2020. – Б.73.
9. Сафаров Ш. Прагмалингвистика. – Тошкент: Ўзбекистон Миллий энциклопедияси, 2008. – Б. 242.
10. Давлатова Р. Ўзбек тилининг дейктик бирликлари: Филол. фан. док. ... дисс. – Тошкент, 2020. – Б. 132.
11. Ключкова Е.С. Повтор как средство создания референциальной структуры гипертекста // Вестник СамГУ, 2008. – № 4 (63) – С. 52.

FRANSUZ VA O'ZBEK MAQOLLARIDA OLAM MANZARASINING AKS ETISHI

Djalilova Difuza Baxodirovna

O'zbekiston Milliy universiteti doktoranti

E-mail: difuza.jalilova2024@gmail.com

Annotasiya: Mazkur maqolada frantsuz tilshunosligida maqol atamasining kommunikativ birlik sifatidagi tavsifi organiladi. Unda frantsuz tilshunosligida maqollarni matal, aforizmlar orasidan aniqlash parametrlari tahlil etiladi. Shuningdek, frantsuz maqollarining baynalminallik belgilari tadqiq etiladi.

Kalit so'zlar: maqol, matal, aforizm, baynalminal maqollar.

Аннотация: В данной статье рассматривается определение термина пословица во французском языкознании как коммуникативной единицы. Анализируются параметры определения пословиц во французском языкознании от пословиц и афоризмов. Также исследуются транснациональные признаки французских пословиц.

Ключевые слова: пословица, предложение поговорка, изречение, интернациональные пословицы.

Annotation: This article discusses the definition of the term proverb in French linguistics as a communicative unit. The parameters of the definition of proverbs in French linguistics from proverbs and aphorisms are analyzed. The transnational features of French proverbs are also researched.

Keywords: proverb, sentence, aphorism, international proverbs.

Fransuz tilshunosligida "maqol" atamasi kommunikativ birliklar va uning tarkibiga kiradi. Boshqa tillar kabi frantsuz tilida bu atama bir necha ma'nolarda ya'ni "proverbe", "sentence", "aphorisme" [2] kabi tushunchalar o'rnida ham ishlatiladi.

"Proverbe" atamasi shakllanishining tarixiy bosqichiga nazar soladigan bo'lsak tilshunos A.Reyning "Dictionnaire de proverbe" kitobida ko'plab ma'nolari keltirilgan [5]. Aflotun o'z davrida bu tushunchani muayyan tuzilishga ega bo'lgan "so'zlashish uslubi" deb talqin qilgan. Suqrot davrida esa maqol xalq orasida keng tarqalgan hikmatli so'z sifatida tushunilgan.

Qadimgi davr va O'rta asrlar maqol va uning mazmunining o'ziga xos xususiyatlari mifologik rivoyatlarga ishora qiladi. Shu bilan birga asosiy so'zni ta'kidlaydigan so'zning maxsus shakli beriladi.

Uygonish davrida maqol faqat intellektual ifoda qilishning usullaridan biri sifatida qabul qilinadi. Shu bilan birga ularning strukturaviy xususiyatlari hisobga olinmaydi. Xristian an'analari ta'siri bilan bogliq bo'lgan axloqiy haqiqatlarga ko'proq nazar soladi.

A.Reyning fikricha, boshqa davrlarda maqolning ta'rifi mavhum tushuncha bo'lgan. Fransiyada klassizm davrida asosiy e'tibor ijtimoiy funksiyalarga va xalqning mentaliteti, xarakteri hamda dunyoqarashini maqolda aks ettirishga qaratilgan. Biroq hozir ham "maqol" atamasi va uning o'xshash atamalar bilan bog'liqligi haqida ko'plab bahsli fikrlar mavjud.

Bir tomondan maqol va matal birliklarining farqlanishi va shu kabi tushunchalar, xuddi "aforizm", "maksima", "sentensiya" kabi tushunchalar qiyin emas, chunki bu holda mualliflik me'zoni qo'llanilishi mumkin. A.Reyning fikricha maqollarning asosiy xususiyati, boshqa frazealogik va paremiologik birliklardan farqli o'laroq xalq orasida yaralishi deb hisoblanadi. Bu fakt esa maqol va matallarda mualliflikni bo'lishini istisno qiladi.

Boshqa tomondan, vaqt o'tishi bilan bu so'zlarning har biri shu qadar mashhur va tez-tez qo'llaniladiki, mualliflik me'zoni yo'qolishi mumkin. Bu frantsuz lug'atlarida ba'zi muallif gaplarining maqol tipidagi birliklar sifatida tasniflanishini tushuntiradi. Masalan "On ne peut contenter tout le monde et son pere" [5].

"Maqol" tushunchasining belgilari. Fransuz tilshunosligida asosiy ma'nosi maqol-matal birligi bo'lgan "maqol" atamasining rivojlanish tarixini kuzatish va uning o'tmishda nimani nazarda tutganini va hozir nimani anglatishini aniqlash maqsadida biz ushbu atamaning fransuz lug'atida qay ma'noda qo'llanilishini tahlil qildik.

"Maqol" atamasining ta'riflarini "France dictionnaire de L'Academie" ning 4- nashrida "maqol" atamasi oz sonli so'zlardan iborat bo'lgan va mashur, ommabop bo'lib qolgan o'ziga xos sentensiya "maksim" sifatida tavsiflanadi".

"Maqol" atamasi ko'p hollarda ko'chma ma'noda ifodalanadigan birlik sifatida berilgan. Shu bilan birga "Dictionnaire de l'Academie francaise"ga ko'ra bu atama shoh Sulaymonning maksumlarini o'z ichiga olgan to'plamning sarlavhasida uchraydi, bu o'sha paytda tilshunoslikda yugoridagi barcha so'z turlarining "On appelle:Proverbes de Solomon" maqollari atamasi bilan bog'liqligini ko'rsatadi [2]

Bu holatda "maqol" atamasini maqol-matal birligining yagona ma'nosida ishlatishning iloji yoqligi shuningdek, "Proverbes de Solomon" to'plamida o'z ta'rkibida axloqiy xarakterdagi qisqa hikoyalarni ifodalovchi parabolalarni o'z ichiga olganligi bilan tasdiqlanadi. Ehtimol bu holda „maqol“ atamasi bilan bir vaqtning o'zida bir nechta ma'nolarda, shu jumadan parabolalarni bog'lash mumkin masal ma'nosida qo'llaniladi. Bu haqiqatni rus tilida nashr etilganligi bilan tasdiqlash mumkin. "Shoh Sulaymonning masallari kitobi" nomi kitobining inglizcha nashri ("The parables of Solomon") nomi ostida nashr etilgan bo'lib, boshqa tillardagi parabola va masal atamaları o'rtasidagi munosabatni ko'rsatadi.

"Maqol" atamasining uchinchi sharhi uning qisqacha o'yin ma'nosi bilan boshlanishining kelib chiqishiga ishora qiladi. O'yinning ma'nosi sahnani o'z-o'zidan ijro etish edi (ehtimol kulgili xarakterga ega) uning maqsadi ma'lum bir maqolning tebranishini tomoshabinlar tahmin qiladigan tarzda yetkazish edi.

Biz ko'plab maqollarni o'zimizniki deb hisoblaymiz ammo uning tub ildizi fransuz tiliga borib taqaladi. Masalan "hayot shunday!", "ayol izla" kabi maqollar "ketish-o'lish oz" degan maqollar (Edmond Arokour 1856-1941 "Rondel de l'Adieu" she'ridan olingan ibora oxir-oqibat maqolga aylangan, o'zbek tilida Ishtaha ovqat bilan birga keladi" va boshqalar [3].

Bazi maqollar umum e'tirof etilgan, ya'ni butun dunyoda ma'lum va mashur ammo turli tillarda ular mos ravishda boshqacha tushuniladi shuning uchun fransuz maqollarining ruscha ekvivalentlari bor va faqat shu millatga xos bo'lgan so'zlarning variantlari mavjud. Biz shu millatlarga xos bo'lgan boshqa variantlarini ko'rib chiqamiz. Misol uchun "À la guerre comme à la guerre".

O'zbek tilida: Urushdagi kabi. "Ma vie, Mes Regles". O'zbek tilida: mening hayotim qoidalarimga muvofiq. "Il faut manger pour survivre, et non pas survivre pour manger" O'zbek tilida: Yeyish uchunmas yashamoq uchun yeyish kerakdir va boshqalarni misol keltirish mumkin.

Maqollar, matallar va aforizmlar ko'plab millatlar madaniyatining ajralmas qismidir. Maqol va matallar yordamida xalq hikmatlari va aforizmlar yordamida xulosalar ifodalanadi. Maqol va matallarning so'zlarini va kichik grammatik tuilmalarini o'rganish ham muhim ahamiyatga egadir. Quyidagi magollarni ko'rib chiqish mumkin:

"Ami de tous, ami de personne" (Har kimning do'sti hech kimga do'st bo'lmas), "Si tu veux être un bon écrivain, écris" (Yaxshi shoir bo'lishni istasangiz ko'proq yozing), "Qui ne edit mot"(Sukut rizo alomati), "L'argent ne fait pas le bonneur"(Baxtni sotib olib bo'lmas pulga), Mashhur tilshunos Dal aytganidek, "Maqollar to'plami-bu xalq tilidan tajribadan olingan hikmatlar majmuasi, sog'lom aql sarasi, xalqning hayotda ortirgan haqiqati" [4]. Turli tillarning magollariga to'xtalar ekanmiz ular xalqning o'sha tarixiy, ma'naviy va moddiy madaniyati kuzguzi ekanligiga guvoh bo'lamiz. Shu sababli turli til maqollarini qiyosiy urganish o'sha xalqning o'ziga xos madaniy va milliy qirralarini ochishga yordam beradi, va boshqacha aytganda, o'sha xalqning mentalitetini ko'rsatadi. Demak, xalq madaniyatini ko'rsatishda maqollar eng faol vositalardan biri

hisoblanadi. Teliya - maqollar butun xalqning madaniyatini ko'rsatuvchi omil ekanligini ishonadi. Har bir maqol uning fikricha xalqning turmush tarzini qisqa va lo'nda umuman olganda to'laligicha ifodalay oladigan ko'zgusi hisoblanadi. Maqollarning linvokultulogik sifatlariga to'xtalgan holda olimlardan Nida maqollarni o'rganishda bevosita tilnin va usha xalqning madaniyatini o'rganish tabiiy jarayon deb hisoblaydi. Tilda mavjud bo'lgan leksima o'sha millatni turmush tarzida kelib chiqqan bo'lib bu bevoita maqolarda ishtirok etadi va maqollar orqali xalqning madaniyati aks etadi. Fransuzcha: "Ange à l'eglise et diable à la maison" maqolini uzbek tiliga tarjima qilsak-cherkovda farishta, uyda shaytondek ma'nosini beradi. Bu o'zbek tilida: Ko'cha-xandon, uy-zindon maqoli orqali ifodalanadi. Agar maqollarning ma'no mazmunini tahlil qilsak ularning fikricha doimiy ko'chada hushchaqchaq xursand bo'lib ko'ringan shaxs, aslida uning teskarisi ya'ni insoniy sifatlariga ega bo'lmaganligini ko'rsatayabdi. Biz yana maqol orqali cherkov-Xristian olamining muqaddas dargohi ekanligi ko'rsatsa Islom olamida esa masjid shu o'rinda ekanligini ko'rishimiz mumkin.

Biz fransuz maqollariga to'xtalar ekanmiz, fransuz xalqining xarakteri haqida so'z yuritib o'tishimiz lozimdir. Dunyo millatlari orasida ushbu xalq vakillari sermulozamatligi va hurmattalabligi bilan ajralib turadi. Bu jihat esa fransuz maqollarida xam o'z aksini topadi:fransuzcha:Il ne fout pas jamais dire:Fontaine,je ne boirai pas de ton eau. O'zbekcha tarjimasi:Hech qachon, favvora, men sening suvingdan ichmayman dema. Buni tahlil qilsak hech qachon kimnidir kamsitmaslik, pastga urmaslik insonlarga nisbatan doimo hushmuomila bo'lish, har qanday qiyin vaziyatda xam insoniylik chegarasidan chiqmaslik haqida aytiladi. Ya'ni maqoldan har ikki millatda ham sermulozamatlilik, ochiqko'ngillik kabi insoniy fazilatlar borligini, xushmuomila bo'lish esa huddi biz singari fransuzlar mentalitetiga xos bo'lgan tushunchalar ekanligini ko'rishimiz mumkin.

Xulosa qilib aytganda, maqollar bu xalqning madaniy merosidir. Ularda o'sha xalqning barcha o'y-fikrlari, turmush tarzi, dunyoqarashi, fe'l-atvori va e'tiqodi aks etadi. Har bir millat o'ziga xos tavsiflarga ega ekan, bu ularning maqollariga ham ta'sir etmay qolmaydi.

Foydalanilgan adabiyotlar:

1. Толковый словарь живого великорусского языка: [в 4 т./[соч.] Владимира Даля. - 3-е изд., испр. и знач. доп., изд. под ред. [и с предисл.] проф. И.А. Бодуэна-де-Куртенэ. Т. 1-4. - Санкт-Петербург; Москва: т-во М. О. Вольф, 1903-1911. - С.15.
2. Dictionnaire de l'Academie francaise. Hachette. Paris: 1989 - P.191.
3. La Fontain. Le Meunier son Fils et l'Ane. Paris -1668-P.100
4. Montreynaud, Pierron, Suzzoni, 2011, P10. Hachette. Paris 1989 - P.220
5. Rey A. Dictionnaire de proverbe. Hachette. Paris 1989 - P.105.