

**O'ZBEKISTON RESPUBLIKASI RAQAMLI
TEKNOLOGIYALAR VAZIRLIGI OLIY TA'LIM, FAN VA
INNOVATSIYALAR VAZIRLIGI
MIRZO ULUG'BEK NOMIDAGI O'ZBEKISTON MILLIY
UNIVERSITETI**



RAQAMLI TEKNOLOGIYALAR VA
SUN'IY INTELLEKTNI RIVOJLANTIRISH
ILMIY-TADQIQOT INSTITUTI



**«KOMPYUTER LINGVISTIKASINING
ZAMONAVIY TEKNOLOGIYALARI – CTCL .2023»**

**VAZIRLIK MIQYOSIDAGI ILMIY-NAZARIY ANJUMANI
ILMIY MAQOLALARI VA TEZISLARI
TO'PLAMI
(2023-YIL 14-APREL)**



TOSHKENT – 2023

KOMPYUTER LINGVISTIKASINING ZAMONAVIY TEXNOLOGIYALARI.
Vazirlik miqyosidagi ilmiy-nazariy anjumani ilmiy maqolalari va tezislari to'plami (2023-yil 14-aprel). –T.: UzMU, 2023. – 288 bet.

Mazkur ilmiy-amaliy anjuman maqolalari va tezislari to'plamida kompyuter lingvistikasi korpus lingvistikasining dolzarb masalalari, 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 kabi masalalarga bag'ishlangan tezislardan tashkil topgan.

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

СОВРЕМЕННЫЕ ТЕХНОЛОГИИ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ.
Сборник научных статей и тезисов научно-теоретической конференции (14 апреля 2023 г.). – Т.: НУУ, 2023. – 288 с.

В сборник материалов научно-прикладной конференции вошли статьи и тезисы по актуальным дисциплинам компьютерной лингвистики, а именно: корпусной лингвистике, компьютерным технологиям в языковом образовании, анализу языка и речи в НЛП (Лингвистический анализатор: морфологический, синтаксический и семантический анализ, анализ и синтез речи), лингвистика и лингвистика программ машинного перевода, компьютерная лексикография: онтология, тезаурус.

Материалы научно-прикладной конференции предназначены для профессоров, преподавателей, докторантов, независимых исследователей, магистрантов и бакалавров, работающих в области компьютерной лингвистики, компьютерных технологий, прикладной филологии.

Mas'ul muharrir:

f.f.d., prof.v.b. N. Abduraxmonova.

Ответственный редактор:

д.ф.н., и.о.проф. Н.Абдурахмонова

Tahrir hay'ati:

F.f.n., prof. Z. Tahirov;

F.f.n., dots. D. Fattaxova;

Katta o'qit. N. Inogamova;

Katta o'qit. N. Bekmuxamedova;

O'qit. F. Topildiyeva;

O'qit. O.Yulbarsov;

O'qit. F.Aliyev;

O'qit. M.Jo'raqulova.

Редакционная коллегия:

к.ф.н., проф. Тахиров З.;

к.ф.н., доц. Фаттахова Д.;

препод. Инагомова Н.;

препод. Бекмухамедова Н.;

препод. Топилдиева Ф.;

препод. Юлбарсов О.;

препод. Алиев Ф.;

препод. Журакулова М.

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

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

MUNDARIJA

1-SHO‘BA

KORPUS LINGVISTIKASINING DOLZARB MASALALARI

N.Z. ABDURAXMONOVA, N.B. TULYAGANOVA. TA‘LIMIY KORPUS YARATISH TAJIRIBASIDAN	7
S.A. KARIMOV, M.S. TURSUNOV. UZBEKCORPORA.UZ: O‘ZBEK TILI MILLIY KORPUSIDA MATN BILAN ISHLASH	13
G.II. TOIROVA. СОЗДАНИЕ И ЗНАЧЕНИЕ ЯЗЫКА КОРПУС В УЗБЕКИСТАНЕ	22
N.Z. ABDURAXMONOVA, K. RAJABOVA, S. XASANOV, D. JUMANAZAROVA, D. O‘ROQOVA, R. YO‘LDOSHEVA. OG‘ZAKI NUTQ KOMPYUTER KORPUSLARI	27
M.S. SHARIPOV, L.U. KURBANOVA, R.U. QURBANOVA. O‘ZBEK TILI KORPUSI DASTURIY TA‘MINOTINI YARATISH	33
M.A. ABJALOVA. PARALLEL KORPUSLARNI YARATISH ASOSLARI	37
G.X. NURMATOVA. DISCUSSION OF NANO-PREFIXED TERMS	44
N. TOSHPO‘LATOVA. KORPUS LINGVISTIKASINING PROGRESSIV MASALALARI	48
G.H. XIDIPOVA. О НЕОБХОДИМОСТИ СОЗДАНИЯ НАЦИОНАЛЬНОГО КОРПУСА УЗБЕКСКОГО ЯЗЫКА	52
K.Q. VASLIDDINOVA. KORPUS LINGVISTIKASI VA DUNYO TILSHUNOSLIGIDA KORPUSGA OID NAZARIYA VA TADQIQOTLAR	57
O.O. YULBARSOV, Z.H. RUSTAMOVA. KORPUS–TABIY TILLARNING MILLIY XUSUSIYATINI SAQLAB QOLISHDA MUHIM INSTRUMENT	61
Z. ISMADIYOROV, T. KAMILOV. SEMANTIK ANNOTATSIYALANGAN KORPUSLAR TAVSIFI	65
M.K. SADIKOVA. KOMMUNIKATIV TIL IMKONIYATLARINI RIVOJLANTIRISHDA KORPUS TEXNOLOGIYALARIDAN FOYDALANISH	68
G. ARABBOYEVA. USING DIGITAL METADATA IN ELECTRONIC CORPORA	72
I.I. ANORBOYEVA, M. JO‘RAQULOVA. PARALLEL KORPUSNING O‘RGANILISHIGA DOIR TAJIRIBALAR TAHLILI	77
D.K. TOSHPO‘LATOVA. ILMIY USLUBDA SOHAVIY TERMINOLOGIYANI YARATISHDA VALENTLIK TUSHUNCHASI	82
S.I. QO‘ZIYEVA. O‘ZBEK KORPUS LINGVISTIKASIDA XI-XIX ASR ADABIY MANBALARI LEKSIKOGRAFIYASINING O‘RNI	89
K. MUKHAMMADIYEVA, M. MUKHAMMADJONOVA, D. RAVSHANOVA, N. SHARIPOVA. UZBEK ELECTRONIC CORPUS OF NEWSPAPER TEXTS	92
T.R. YANDASHOVA. INGLIZ VA O‘ZBEK TILI FRAZEMALARIDA INSON QIYOFASI VA XULQ-ATVORI IFODASI	96

A.SH. ISMAILOV, G.T. QODIROVA, M.A. YIGITALIYEVA, R. XUSANOVA. O'ZBEK STEMASI ALGORITMI	101
M.X. ESHMAMATOVA. KORPUS LINGVISTIK TA'MINOTINI YARATISH MASALALARI	106

2-SHO'BA

TIL TA'LIMIDA KOMPYUTER TEXNOLOGIYALARI

G.S. KELDIYOROVA. O'ZGA TILLARNI O'QITISHDA INTERNETNING O'RNI	110
A. TESHABOYEV, M.Z. SAYFUTDINOVA. CHET TILLARINI O'QITISHDA AXBOROT TEXNOLOGIYALARINING O'RNI	114
N. BEKMUXAMEDOVA. O'ZGA TILLI GURUHLARDA TIL O'QITISHDA MULTIMEDIA TAQDIMOTIDAN FOYDALANISH	118
F.R. TOPILDIYEVA. TA'LIM JARAYONIDA KOMPYUTER TEXNOLOGIYALARI, MULTIMEDIALI O'QUV KONTENTLARIDAN FOYDALANISH O'RNI VA SAMARASI.....	121
D.A. FATTAHOVA. TILLARNI O'QITISH JARAYONIDA MULTIMEDIALI O'QUV KONTENTLARIDAN FOYDALANISH	125
Б.Х. ДАНИЯРОВ. АКАДЕМИК ЁЗУВ ДАРСЛАРИДА ЗАМОНАВИЙ СМАРТФОНЛАРДАН САМАРАЛИ ФОЙДАЛАНИШ	135
X.X. AXMEDOVA, D.A. FATTAHOVA. ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННО КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ ПРИ ОБУЧЕНИИ ЯЗЫКУ	138
Н.Ш. ИНОГАМОВА. ИСПОЛЬЗОВАНИЕ СОВРЕМЕННЫХ ИКТ В ПРОЦЕССЕ ОБУЧЕНИЯ ЯЗЫКАМ	143
S.B. BOYSARIYEVA. TIL O'RGANISHDA KYTO' VA MYTO'DAN FOYDALANISH	147
Z.SH. ABDULLAYEVA, G.A.SHIHNAZAROVA. INTELLEKTUAL TAHLIL ASOSIDA CHET TILI BILIMINI TEKSHIRISHNI MODELLASHTIRISH	150

3-SHO'BA

NLPDA TIL VA NUTQ TAHLILI (LINGVISTIK ANALIZATOR: MORFOLOGIK, SINTAKTIK VA SEMANTIK TAHLIL, NUTQ ANALIZI VA SINTEZI)

E. NAZIROVA, SH. ABIDOVA, M. UZAKOVA. TURKIY TILLAR O'RTASIDA TABIIY TILLARNI QAYTA ISHLASH MODEL (O'ZBEK TILI MISOLIDA)	153
Л. БОБОЕВ. МАТНЛИ ХУЖЖАТЛАРНИ ТАСНИФЛАШНИНГ ГИБРИД АЛГОРИТМИНИ ИШЛАБ ЧИҚИШ	159
Н. РАВШАНОВ, И.И. БАКАЕВ. ПРОЕКТИРОВАНИЕ MORFOLOGICHESKOGO ANALIZATORA DLYA ESTESTVENNYKH YAZYKOV NA PRIMERE UZBEKSKOGO YAZYKA.....	163
U. TULIYEV. ANALYSIS OF THE SEMANTIC RELATEDNESS OF WORDS BASED ON A COMMON DICTIONARY	172
M.YA. URAZALIYEVA. KORPUSDA MORFOLOGIK ANALIZATOR UCHUN LINGVISTIK MODELLASHTIRISH MASALALARI.....	176
A. SHAYMARDANOV. RECENT ADVANCEMENTS IN UZBEK NLP TOOLS.....	181

S. ISRAILOVA, M. ERGASHEVA. BILINGVIZM TURLARI VA IKKINCHI TIL LEKSIK BAZASI	185
YU.B. BAHRIDDINOVA. KOMPYUTER LINGVISTIKASINING NUTQ BIRLIKLARI ANALIZATORLARI VA ULARNING TAHLILI	189
G. SHAMSIYEVA. O'ZBEK-INGLIZ TILLARI UCHUN PARALLEL KORPUSGA ASOSLANGAN MASHINA TARJIMASI TEXNOLOGIYASI.....	192
F.T. ALIJONOVA. TABIIY TILNI QAYTA ISHLASH JARAYONIDA YUZAGA KELADIGAN ASOSIY MUAMMOLAR	196
X.K. RAXIMOV. SENTIMENT TAHLILGA OID XORIJ TADQIQOTLARI	202
Z.T. TAXIROV. TIL VA USLUB MATN TAHRIRINING BIRINCHI BOSQICHI SIFATIDA.....	206
X.I. AXMEDOVA. IMPORTANCE OF STATISTICAL METHODS IN DETERMINE THE HOMONYMY	210
G'R. MATLATIPOV, M.SH. OTAMURATOV, M. G. RAMATOVA. O'ZBEK TILSHUNOSLIGI UCHUN QISQARTIRILMAYDIGAN GRAMMATIKA TUZISH.	216
SH. TO'RAYEV. O'ZBEK TILI MATNIDAGI NOSTANDART SO'ZLARNI NORMALLASHTIRISH MUAMMOLARINI YECHISH.....	220
A.K. AVEZOVA. MASHINA TARJIMASIDA ANTROPONIMLAR BILAN ISHLASH (O'ZBEK VA TURK TILLARIDAGI ANTROPONIMLAR MISOLIDA).....	225
SH.A. ABDURAXMONOVA. BOLALARNING MATN PERSEPSIYASIDA SO'ZLARNING ESLAB QOLINISHI	229

4-SHO'BA

KOMPYUTER LEKSIKOGRAFIYASI: ONTOLOGIYA, TEZAURUS

U. QO'ZIYEV. MA'LUMOTLAR BAZASINI SHAKLLANTIRISHDA SO'ZLARNI TASNIFLASH MUAMMOLARI.....	234
S.T. ISRAILOVA. TIL BIRLIKLARNING SEMANTIK XUSUSIYATLARI.....	238
A. OTEMISOV. TURKIY TILLARDA CHASTOTALI LUG'AT TUZISH TARIXIDAN	242
Q. XOSHIMJONOVA. O'ZBEK TILI TEZAURUSI UCHUN OT SO'Z TURKUMI DOIRASIDAGI AYRIM KONSEPTLARNING PSIXOLINGVISTIK TADQIQI.....	245
Z.G. QODIROVA. BILIMLAR BAZASINI ONTOLOGIK MODELLARDA IFODALASH	250
G. LATIPOVA. O'ZBEK TILI VALENTLIK LUG'ATINI YARATISHDA KORPUSGA ASOSLANGAN TAHLIL	255
M.F. SAPAROVA. THE BASIC FEATURES OF THESAURUS DICTIONARIES.....	259
S.M. XUDOYBERGANOV. VERBS IN THESAURUS	263
M.S. JO'RAQULOVA, G. XOLIYOROVA. KOMPYUTER LEKSIKOGRAFIYASIDA ASSOTSIATIV MAYDON BIRLIKLARINING GENDEROLOGIK TADQIQI.....	266
L.J. BABAJANOVA. TILSHUNOSLIK AXBOROT TIZIMIDA TEZAURUSDAN FOYDALANISH ASOSLARI.....	271

S.A. TURSUNOVA. KOMPYUTER LEKSIKOGRAFIYASI – O‘ZBEK TILSHUNOSLIGINING MUHIM YUTUQLARIDAN BIRI SIFATIDA	276
M. KARIMOVA. RUS TILSHUNOSLIGIDA ALLUZIYA: O‘RGANILISHI VA TURLARI.....	280
S. OYBEKOVA. BILINGVAL TEZAURUS YARATISHDA SOHAVIY TERMINOLOGIYA MASALALARI	283
D. MURODOVA. RUSCHA-O‘ZBEKCHA MASHINA TARJIMASIDA MATN TAHLILI MASALALARI	287

1-SHO‘BA KORPUS LINGVISTIKASINING DOLZARB MASALALARI

TA'LIMIY KORPUS YARATISH TAJRIBASIDAN

Abduraxmonova Nilufar Zaynobiddin qizi,

O‘zbekiston Milliy universiteti professori,
filologiya fanlari doktori,

E-mail: n.abduraxmonova@nuu.uz

Tulyaganova Nilufar Baxtiyorovna,

O‘zbekiston Milliy universiteti o‘qituvchisi

Annotatsiya: Ushbu tezisda korpusning ta’limiy ahamiyati va buroda yaratilgan korpus namunalari tahlili keltirilgan. Ta’limiy korpuslar ingliz tili misolida juda keng qamrovli ishlarni o‘z ichiga oladi. Ularga tilni ona tili sifatida, ikkinchi va xorijiy til sifatida o‘qitish kabi mezonlar ishlab chiqilganligi misol bo‘lishi mumkin. Korpus kontentiga biriktirilgan matnlar yoki og‘zaki nutq namunalari (audio yoki multimedia fayllar) foydalanuvchi maqsadi uchun moslashtirilishi, qolaversa, korpusdan foydalanish qulaylikka ega bo‘lishda metama’lumotlarni berish me’yorlariga amal qilishning muhim jihatlari tezisda o‘z ifodasini topgan.

Kalit so‘zlar: ta’limiy korpus, xalqaro korpus, korpusga asoslangan tahlil

Annotation: The paper presents the didactic significance of the corpus and the analysis of corpus samples created in it. Learner corpora include a very wide range of works, as in the case of the English language. An example is the development of criteria such as teaching the language as a mother tongue, as a second language, and as a foreign language. Adjustment of texts or speech samples (audio or multimedia files) attached to the corpus content for the purpose of the user, as well as the essential aspects of following the norms of providing metadata to make the corpus easier to use are expressed in the article.

Keywords: learner corpus, international corpus, corpus analysis.

Аннотация: В статье представлена дидактическая значимость корпуса и анализ созданных в нем корпусных образцов. Учебные корпуса включают в себя очень широкий спектр работ, как и в случае с английским языком. Примером может служить разработка таких критериев, как преподавание языка как родного, как второго языка и как иностранного языка. В статье изложены корректировка текстов или речевых образцов (аудио или мультимедийных файлов), прикрепленных к содержанию корпуса, для целей пользователя, а также существенные аспекты соблюдения норм предоставления метаданных для облегчения использования корпуса.

Ключевые слова: образовательный корпус, интернациональный корпус, корпусный анализ.

Til va tilshunoslikni o‘rganilishida korpus manba sifatida foydalanish bo‘yicha omma qiziqishining ortib borishi ICAME (the international computer archive of modern and medieval

English) a'zolari o'rtasida bo'lib o'tgan muhokama natijasida 1994- yil TaLC yaratilishiga sabab bo'ladi. Dastlabki ta'limiy korpusga oid konferensiya 1994- yilda Lankasterda bo'lib, tadqiqotda kompyuter matn korpuslaridan foydalanish yaxshi yo'lga qo'yilgani va ular o'qitish maqsadlarida keng foydalanib kelinyotganligini, korpus ma'lumotlaridan o'quv materiallarini tayyorlashda hamda xorijiy tillarni va lingvistik ta'limda talabalar tomonidan bevosita korpuslarning yaratilishiga doir muhokmalar kun tartibiga qo'yiladi.

Dastlabki ta'limiy korpuslarni yaratilishi Sylviane Granger va uning ilmiy jamoasi tomonidan "Ingliz tili o'rganuvchilari uchun xalqaro korpus" (ICLE – International corpus of learner English) yaratilgan. Ushbu korpusning maqsadi til o'rganuvchilarining umumiy yuzaga keluvchi xatolarini tasniflash orqali ularni tuzatish metodlarini yaratishga qaratilgan [D. Stewart: 2004].

Ushbu ta'limiy korpusni yaratishda ingliz tili o'rganuvchilarning quyidagi belgilari hisobga olingan [G. Sylviane; 199]: ta'lim muhiti, yosh, ona tilisi, ta'lim bosqichi, vazifalarning tabiati.

Ushbu korpus uchun quyidagi kriteriyalari doirasida til o'rganuvchilar tanlab olingan:

- 1) o'rganuvchilar turi;
- 2) yoshi: yoshi kattalar (adults)
- 3) ta'lim bosqichi: yuqori daraja (advanced);
- 4) vazifasi: esselar

Ingliz tilining ta'limiy xalqaro korpusi ingliz tilining ikkinchi til sifatida fransuz, nemis, xitoy, yapon, chex kabi millatga tegishli o'rganuvchilar asos qilib olingan.

Ingliz tili xalqaro ta'limiy korpusi bir nechta lingvistik instrumentariylardan tarkib topgan.

- The ICE Markup Assistant – ushbu instrumentariy razmetkalash uchun standart belgilar to'plami asosida jarayonni soddalashtirilgan avtomat tizimi sanaladi;
- TAGSELECT – the Ice tag Selection system (teglash tizimi) korpusdagi har bir so'z uchun TOSCA teggeri yordamida generatsiya qilingan so'z guruhlari teglarini avtomatik saralash tizimi sanaladi;
- ICEMARK – TOSCA parseridan oldin teglangan matnni sintaktik razmetkalash uchun qo'llanadi. Shuningdek, boshqa funksiyalaridan biri so'z guruhlari teglarini o'zgartirish, matnda uchragan so'zlarni xatolari tahrirlash, teglanmagan gaplarni ko'rish kabi imkoniyatga ham egadir.
- ICECUP – korpus Utility dasturi ushbu instrumentariy korpusni tahlil qilish, konkordanslash va qidiruv kabu funksional jihatdan ahamiyatga ega.

XIX asrning oxiri va XX asrning boshlarida so'zning klassik nazariyasi iki yirik leksik tadqiqotning yaratilishida katta rol o'ynadi. Roget tezaurus (1852) va Oksford ingliz lug'ati (Murray 1884-1928). Matnlarga qarab so'zlarning giponimi, sinonimi, antonimi va ularning ma'nolari leksik munosabatlarga ko'ra lug'atda beriladi. Har bir so'z bundan tashqari bir yoki bir nechta grammatik kategoriyalarga tegishli bo'ladi va ularning talaffuzi, etimologiyasi va tarixi beriladi. Ushbu nazariyaga ko'ra leksik birlik boshqa tizimlarga ham aloqador bo'lib, grammatika va fonologiya o'rtasida o'zaro munosabat yotadi. Sistem-funksional lingvistikaning markazida ham leksis turadi, zero bu modeldagi tizimlar leksik tanlov eng oxirgisi bo'lganidek ba'zan uni shu tarzda hosil qiladi [D. Stewart:2004]. XIX asrda yaratilgan Tezaurus lug'ati ta'sirida ko'pgina nazariy tadqiqotlar jonlana boshlaydi. XX asrda korpusga asoslangan o'quv lug'atlari Collins Cobuild, Oksford, Longman, Makmillan kabi lug'atlar yartailganligini guvohi bo'lamiz.

Ta'limiy korpuslar 1990-yillarda ommalashgan bo'lsa-da, biroq 1960-70-yillarda tillarni o'rganishda "xatoni tahlil qilish" (error analysis) tarzda metodologik yondashuv sifatida o'rganilgan.

Ma'lumki, so'zlar muayyan darajada kombinatsiya hosil qilish uchun ma'lum turdagi matnlarda qo'llanadi. Bu an'anaviy tilshunoslida so'zlarning leksik ma'no qurshivini ko'rsatishda matn, gap yoki kontekst jumalari bilan qo'llanganini ko'ramiz. Endilikda so'zni bir vaqtda turli maqsadlardan kelib chiqib tahlil qilish imkoniyatini yaratib beruvchi lingvistik instrumentariy sifatida korpusni ta'kidlash o'rinli bo'lar edi. So'zning kollokatsiyalar bilan (so'z birikma yoki turg'un birikma) birgalikda uning tarkibida qo'llanganda so'zga doir u yoki bu ma'no anglashiladi. Demak ayrim so'zlar uchun uning grammatik kategoriyaga xoslanishi gap tarkibida kelganda aniqlashadi: *o'qish* (ot, harakat nomi yoki fe'lining buyruq shakli), *qil* (ot, fe'l), *yosh* (sifat, ot) kabilar.

Qachonki til foydalanuvchisi uchun biror so'z semantik to'plam yoki guruh bilan birgalikda kelganda semantik assotsiatsiyani aniqlash mumkin bo'ladi [D. Stewart: 2004].

Ta'limiy korpusga asoslangan til o'qitish yoki o'rganish jarayoni statistik yoki chastotaga asoslangaligi bilan ahamiyatlidir. Ayniqsa, u yoki bu so'zning kontekstda birikma tarzda so'z qurshovida kela olish imkoniyati so'zlarning ko'p ma'noli yoki omonim bo'lgan holatda juda muhim ahamiyat kasb etadi.

Ta'limiy korpus til o'rganuvchilar tomonidan tayyorlangan matnlarning tizimli kompyuterlashtirilgan majmuasi sifatida ta'rif keltiriladi [N. Nesselhauf; 2005]. Tizimli deganda, muayyan til o'rganuvchilar guruhi sifatida tanlab olingan til bilish darajasiga ko'ra jamlangan matnlardan tuzilgan korpus nazarda tutiladi. Ona tili korpusiga ixtisoslashgan korpusda matnning tabiiy yaratilganligi va tayyor olingan matnlar resurs bo'lib hizmat qilishi bilan ajralib turadi. Ta'limiy korpusda xorijiy til yoki ikkinchi til sifatida o'rganuvchilar uchun mo'ljallangan bo'lib, nazoratdagi til o'rganuvchilar tomonidan yaratilgan chekli matnlardan tashkil topadi. Uar janr jihatdan esselar, og'zaki intervyular bo'lishi mumkin.

Rus tili ta'limiy korpusi rus tilida gapiruvchi turli millat vakillari tomonidan tayyorlangan materiallar asosida yaratilgan.

Korpusdagi matn va muallifga tegishli metama'lumotlar quyidagi ikki maydonga ajratilgan [<http://web-corpora.net/RLC>]:

I. Asosiy maydon:

- Og'zaki / yozma matn
- Xorijiy / ona tili sifatida rus tili
- Muallifning dominant tili
- Rus tilini bilish darajasi

II. Ixtiyoriy maydon:

- Muallif jinsi
- Matn yaratilgan sanasi
- Janri

Ushbu korpus iqtisodiyot oliy maktab korpus texnologiyalari lingvistik laboratoriyasida xalqaro loyiha doirasida yaratilgan. Korpus 730000 token dan iborat bo'lib, matnlarning 56 foizi rus tili ikkinchi til sifatida so'zlovchilar nutqidan, 44 foizi mahalliy (universitet va kollej talabalari) yuqori va o'rta til bilish darajasiga ega rusiy zabon vakillar tomonidan olingan materiallarga

asoslanilgan. Korpusning dastlabki versiyasi faqat Amerika ingliz tilida ona tilisi bo'lgan loyihada cheklangan ishtirokchilar tomonidan yozib olingan. RULEC deb nomlangan subkorpus tarkibida akademik yozishmalardan iborat. Barcha ishtirokchilar ononim shaklda qatnashgan. Boshqa respondentlar ID kod bilan belgilangan [E. Rakhilina: 2016].

Lingvistik jihatdan aniqlangan xatolarni teglar tizimida ko'rsatilgan:

Til sathi	TEG	Izohi
Imloviy xato	Graph	Lotin alifbosidan foydalanish
	Hyphen	Chiziqchani qo'llash bilan xatolik
	Space	So'zlar o'rtasidagi bo'shliqning tushib qolishi
	Translit	Atoqli otlarni transliteratsiya qilishdagi xatolik
	Ortho	Xato yozilgan harf
	Misspel	Bir tokendagi bir nechta o'rinda yo'l qo'yilgan imloviy xatolik
Morfologik xato	Infl	Grammatik qo'shimchlar xato qo'llangan so'z
	Deriv	Yasalgan so'z
	Altern	So'z o'zagini qo'llashdagi xato
	Num	So'zdagi son kategoriyasini qo'llash bilan bog'liq xatolik
	Gender	Jins bilan bog'liq xatolik
	Morph	Boshqa morfologik xato
Sintaktik xato	AgrCase	Kelishik bilan bog'liq xatolik
	AgrGender	Jinsning o'zaro moslashish bilan bog'liq xato
	AgrNum	Son bilan moslashishga doir xatolik
	AgrPers	Shaxs-son bilan bog'liq xatolik
	Asp	Fe'l mayli bog'liq xato
	Passive	Majhul nisbatni qo'llash bilan bog'liq xatolik
	Tense	Fe'l zamoni bilan bog'liq xatolik
	Mode	Fe'lni qo'llash bilan bog'liq xatolik
	Refl	O'zlik nisbatdagi fe'lni qo'llash bilan bog'liq xatolik
	Gov	Xato kelishik
	WO	So'z tartibidagi xatolik
	Ref	O'zlik olmoshi bilan bog'liq xatolik
	Conj	Bog'lovchini noto'g'ri qo'llanishi
	Neg	Bo'lishsizlikning xato qo'llanishi
	Aux	Ko'makchilarni xato qo'llash
	Brev	Sifatlarning qisqa formasini qo'llashdagi xatolik
	Syntax	Boshqa sintaktik xato
	Constr	Gap tuzilishdagi xatolik
	Lex	Leksik xatolik
Leksik xato	CS	Kodlash bilan bog'liq xatolik
	Par	Paronim bilan bog'liq xatolik
	Idiom	Idiom bilan bog'liq xato

Qo'shimcha teglar	Del	So'z yoki biror morfemada belgining tushib qolishi
	Insert	So'z yoki biror morfemada belgining qo'shib qolishi
	Subst	So'z yoki biror morfemada belgining almashib qolishi kelishi
	Transp	So'z yoki biror morfemada belgining noo'rin kelishi
	Transfer	Tildagi kelishikni transfer qilish
	Not-clear	Boshqa noma'lum fragmentlar

Xulosa qilib shuni aytish mumkinki, yuqorida o'rganilgan ta'limiy korpuslarida erishilgan ijobiy natijalar va tajribalardan foydalanib o'zbek tilining elektron korpusida ham ta'limiy korpuslarni yaratish va bub o'yicha metodik ishlanmalarni shakllantirish istiqboldagi ustuvor vazifalardan sanaladi [Abdurakhmonova N.: 2018, 2019, 2020, 2021, 2022,].

Foydalanilgan adabiyotlar:

1. Abdurakhmonova N. Modeling Analytic Forms of Verb in Uzbek as Stage of Morphological Analysis in Machine Translation. Journal of Social Sciences and Humanities Research. 2017;5(03):89-100.
2. 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.
3. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. In Proceedings of the International Conference on Language Technologies for All (LT4All) 2019.
4. D. Stewart, S. Bernardini, G. Aston Introduction: Ten years of TaLC / Corpora and language learners, Vol 17, John Benjamins Publishing Company Amsterdam/Philadelphia, 2004, - P. 3.
5. D. Stewart, S. Bernardini, G. Aston Introduction: Ten years of TaLC / Corpora and language learners, Vol 17, John Benjamins Publishing Company Amsterdam/Philadelphia, 2004, - P. 22.
6. Ekaterina Rakhilina, Anastasia Vyrenkova, Elmira Mustakimova, Alina Ladygina, and Ivan Smirnov. 2016. [Building a learner corpus for Russian](#). In *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition*, pages 66–75, Umeå, Sweden. LiU Electronic Press.
7. <http://web-corpora.net/RLC>
8. 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
9. 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.
10. N. Z. Abdurakhmonova, A. S. Ismailov and D. Mengliev, «Developing NLP Tool for Linguistic Analysis of Turkic Languages,» 2022 IEEE International Multi-Conference on

Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation, 2022, pp. 1790-1793, doi: 10.1109/SIBIRCON56155.2022.10017049.

11. Nadjia Nesselhauf Collocations in a learner corpus John Benjamins Publishing company, Amsterdam, 2005. – P. 40.
12. Sylviane Granger The learner corpus: a revolution in applied linguistics / English Today 39, Vol.10, No.3 (July 1994), Cambridge university press. – P. 26.

UZBEKCORPORA.UZ: O‘ZBEK TILI MILLIY KORPUSIDA MATN BILAN ISHLASH

Karimov Suyun AmirovichFilologiya fanlari doktori, professor
Samarqand davlat universiteti**Tursunov Muhammadsolih Sa’din o‘g‘li**Toshkent axborot texnologiyalari universiteti
Samarqand filiali Mustaqil tadqiqotchisi

Annotatsiya: Ushbu ishda uzbekcorpora.uz o‘zbek tili milliy korpusi uchun matnni tayyorlash va korpusga matnni kiritish to‘g‘risida ma’lumotlar berilgan. Matnlarni korpusga kiritish uchun, uni raqamlashtirish va korpus talabiga ko‘ra formatlashtirish lozim bo‘ladi. Raqamlashtirish jarayonida ko‘proq matnni lingvistik xatolarini tuzatishga e’tibor berish lozim. Bu esa foydalanuvchiga lingvistik ma’lumotlarni yuqori aniqlikda olish imkonini beradi.

Kalit so‘zlar: uzbekcorpora.uz, matn, korpus, matn formati, razmetka, raqamlashtirish.

Annotation: This paper provides information on the preparation of the text for the national corpus of the Uzbek language uzbekcorpora.uz and the inclusion of the text in the corpus. To include texts in the corpus, they must be digitized and formatted in accordance with the requirements of the corpus. In the process of digitization, more attention should be paid to correcting language errors in the text. This allows the user to obtain linguistic information with high accuracy.

Keywords: uzbekcorpora.uz, text, corpus, text format, markup, digitization.

Аннотация: В данной работе представлена информация о подготовке текста для национального корпуса узбекского языка uzbekcorpora.uz и включении текста в корпус. Для включения текстов в корпус необходимо их оцифровать и отформатировать в соответствии с требованиями корпуса. В процессе оцифровки следует уделять больше внимания исправлению языковых ошибок текста. Это позволяет пользователю получать лингвистическую информацию с высокой точностью.

Ключевые слова: uzbekcorpora.uz, текст, корпус, формат текста, разметка, оцифровка.

Korpusda matn bilan ishlash **uch bosqichda** amalga oshiriladi:

- Matnni korpusga tayyorlash bosqichi;
- Matnni korpusga kiritish bosqichi;
- Matn bilan ishlash bosqichi;

Matnni korpusga tayyorlash bosqichi. Raqamli va raqamli bo‘lmagan matnlar korpus uchun matn manbalari sifatida ishlatilishi mumkin. Tabiiyki, ikkinchi holatda, qandaydir tarzda matnni kompyuterga kiritish kerak bo‘ladi: uni qayta yozish yoki skanerlash kerak. Masalan, Alpomish dostonining Fozil Yo‘ldosh variantining docx formati bizda mavjud emas. Shuning uchun, bu kitobni raqamlashtirishimiz kerak bo‘ladi. Bunda qo‘lyozma manbani yoki kitobni

skanerlash jarayonida pdf formatga o'tkaziladi va undagi yozuvlarni tanib oluvchi konvertor (masalan, *Fine Reader*) dasturlar yordamida *Microsoft Office Word* dasturida ochadigan *docx* formatga o'tkaziladi. Bunda albatta konvertor skaner qilingan elektron kitobni *docx* formatga o'tkazishda kitobning asl holati bo'yicha o'tkazishi juda qiyin. Shuning uchun qo'l mehnati asosida *docx* formatdagi matn asl holatdagi matn bilan birxillikka keltiriladi. Bunda imloviy buzilishlar, matndagi belgilarni noto'g'ri tanib olish kabi xatoliklar tuzatiladi [Qarshiyev A.B., Karimov S.A., Tursunov M.S., 2022:58-63].

Biroq hozirgi vaqtda ko'pchilik korpuslar allaqachon raqamli elektron shaklda (kompyuterlashtirishning kuchayishi tufayli) matnlardan iborat. Korpus loyihasi uchun zarur hujjatlarning mavjud elektron versiyalaridan foydalanish mantiqan to'g'ri keladi. Bu korpusni kompilyasiya qilish (dasturni ishga tushirish)ning murakkabligini keskin kamaytiradi.

Raqamlashtirilgan matnlarning aniq manbalaridan biri bu matnlarning ulkan korpusi bo'lgan internetdir. Tarmoqdan istalgan matnni turli xil formatlarda topish imkoniyati mavjud. Lekin internetdan olingan matnni to'g'ridan-to'g'ri korpusga kiritib bo'lmaydi, chunki matn to'g'risida ma'lumotlar yoki matnning mazmuni asl holatdagi matnga nisbatan o'zgartirilgan bo'lishi mumkin [Colm Smyth, 2016:105-109]. Shu sababli korpusga kiritish uchun raqamlashtirilgan matn albatta matnning asl manbasi bilan ustma-ust tushishi zarur. Aks holda, korpus manbasi korpus foydalanuvchilariga noto'g'ri axborot va tushunchalar berishi mumkin.

Korpus matnlari turli xil manbalardan olingan va turli xil formatlarda taqdim etilgan: *oddiy matn*, *html*, *rtf*, *pdf* va boshqalar. Har bir kirish formati uchun tavsiflangan formatlarni *docx* shaklga aylantirishga imkon beradigan konvertorlar va makroslar to'plami yaratiladi. Bunda albatta sintaktik xatolar hosil bo'lishi mumkin. Shu sababli tadqiqotchi matndagi xatolarni to'g'rilab, asl matn bilan bir xillikka olib kelinadi.

Matnni tayyorlashda undan matn muallifi bo'lmagan yoki tilni o'rganish uchun ahamiyatsiz bo'lgan barcha elementlar olib tashlanadi, xususan:

- sarlavhaosti ma'lumotlar, sahifa raqamlari;
- sarlavha sahifalari, tarkibi, chiqish ma'lumotlari, bibliografiya, annotasiya;
- tahririyat izohlari (saqlanib qolingan mualliflik huquqi izohlari);
- chizmalar, diagrammalar, formulalar (lekin ularning ostidagi yozuvlar saqlanib qolgan);
- raqamlarning uzun ketma-ketliklari (jadvallarda).

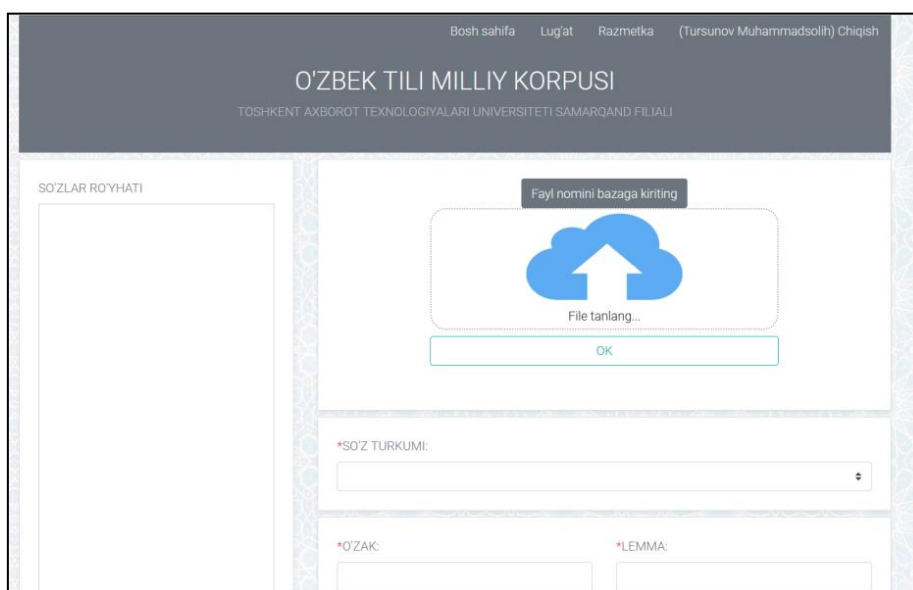
Matnni korpusga kiritish bosqichi. Matn korpusga kiritishga tayyor holatga keltirilganda, uni korpusga yuklash mumkin [McEnery T., Hardie A., 2012:294]. Korpusga matn yuklanish jarayonida *docx* formatdagi matn *json* formatga almashtiriladi va bir vaqtning o'zida matn ham tokenlanadi, ham har bir so'zga grammatik ma'lumotlar biriktiriladi. Grammatik ma'lumotlarni har bir so'zga biriktirishda dastur serverga murojaat qiladi. Serverda so'zlar lug'ati mavjud bo'lib, har so'zning grammatik xarakteristikalarini oldindan berilgan. Bunda matnni razmetkali matnga o'tkazishda ikki bosqichdan o'tadi. Birinchi bosqichda matn tokenlanadi (token – matn tarkibidagi eng kichik birlik (so'z, qisqartmalar, raqamlar va h.k.)). Ikkinchi bosqichda tokenlangan matn razmetkalanadi (tokenlarning lingvistik xususiyatlarini ko'rsatish) [Бочаров В.В., Грановский Д.В., 2011:104-109]. Razmetkalash jarayonida dastur serverga so'zni yuboradi va server shu so'zni grammatik harakteristikasini dasturga javob tariqasida qaytaradi. Va oddiy matnni razmetkalanagan matn holatiga o'tkazadi.

Matnni korpusga kiritishdan avval quyidagi muhim ko'rsatmalarga e'tibor berish lozim:

- Matn fayli *Microsoft Word* dasturida tayyorlangan bo'lishi lozim;

- Faylning formati **.docx* bo'lishi kerak;
- Matn shrifti *Times New Roman* va o'lchami *14 px* bo'lishi tavsiya etiladi;
- Faylda matnning sahifasi katta bo'lsa *10 sahifadan* qilib, fayllarga bo'lib chiqish tavsiya etiladi. Sababi dastur tomonidan matnni avtomatik qayta tayyorlash jarayoni uzoq vaqt davom etishi mumkin.
- Matnning *imloviy xatolarini* to'liq tekshirib, tuzatish kiritish kerak.
- Qo'shma yoki juft so'zning o'rtasida *"-"* belgisi mavjud bo'lsa, bu belgining o'ng va chap tomonida *bo'sh joy (probel)* bo'lmasligi kerak.
- Manba matnning *asl holati* saqlanib qolishi lozim.

Dastur bilan ishlash. Foydalanuvchi shaxsiy kabinetga kirganida quyidagi oyna ochiladi (1-rasm).



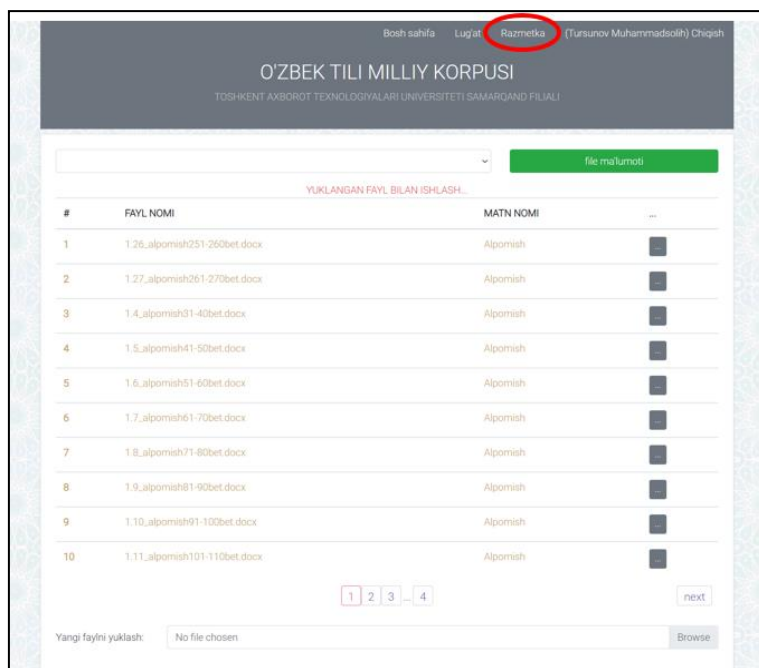
1-rasm. Asosiy oyna.

Razmetka nomli menyu yordamida korpusga matn kiritiladi (2-rasm).



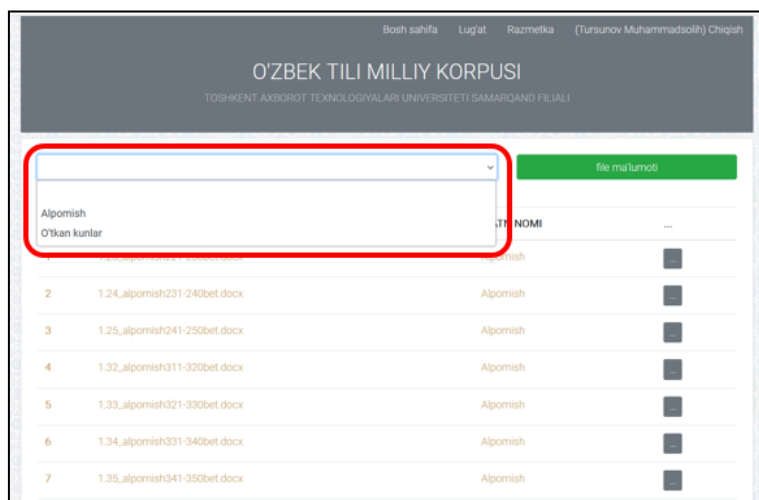
2-rasm. *Matn kiritish* bo'limiga o'tish.

Razmetka bo'limida tadqiqotchi tomonidan korpusga yuklangan matnlar ro'yxati ko'rinadi (3-rasm). 3-rasmida ko'rinib turganidek bir nechta yuklangan fayllar bor. Ushbu ro'yxatdan istalgan birining ustidan sichqonchani tekkizsa, razmetkalanga matn ochiladi.



3-rasm. Razmetka sahifasining ko'rinishi.

Korpusga **matn faylini yuklashdan** avval, ushbu fayl tegishli bo'lgan dokument nomi tanlanadi. (4-rasm).

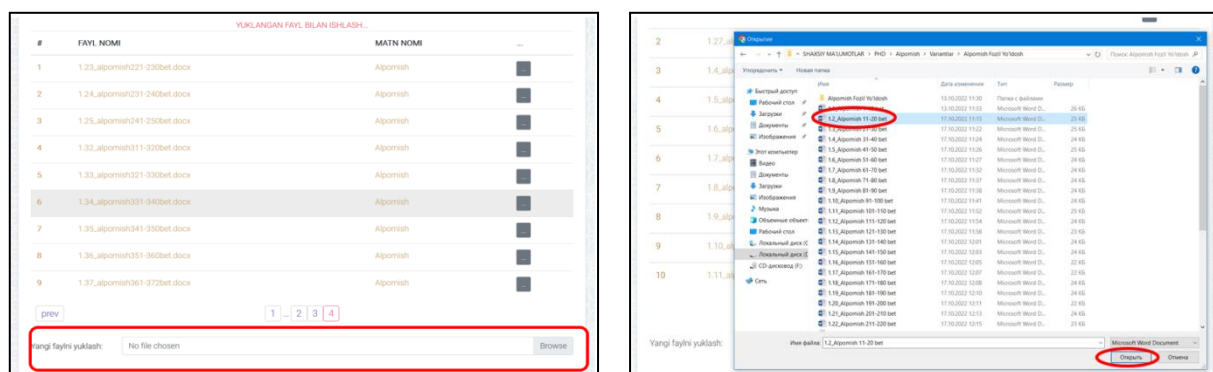


4-rasm. Dokument nomini ko'rsatish.

Agar yuklanishi kerak bo'lgan faylga biriktiriladigan **dokument nomi** ro'yxatda mavjud bo'lmasa, u holda "**Fayl ma'lumoti**" tugmasi bosiladi va ro'yxatga **dokument** to'g'risidagi ma'lumotlar tadqiqotchi tomonidan kiritiladi (5-rasm). Dokument to'g'risida **ma'lumotlar** to'ldiriladi va **OK** tugmasi bosiladi. Undan so'ng dokument nomi ro'yxatida yangi dokument nomi paydo bo'ladi.

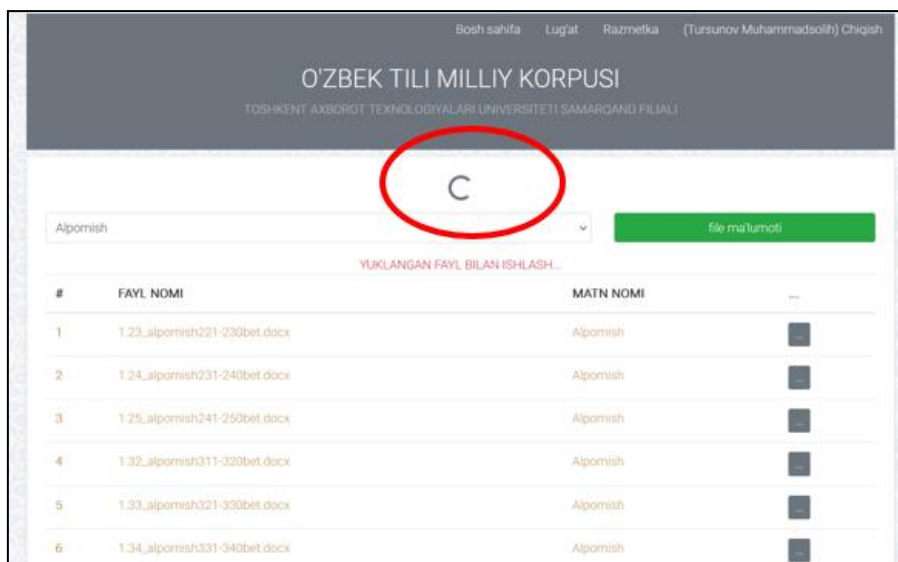
5-rasm. Yangi dokument to'g'risida ma'lumotlarni kiritish.

Dokument nomi ko'rsatilgan so'ng, ko'rsatilgan maydon yordamida yangi faylni **korpusga yuklash** mumkin. Korpus uchun tayyorlangan **matn tanlanib**, “Открыть” tugmasi bosiladi (6-rasm).



6-rasm. Korpusga matn yuklash.

Faylni yuklash jarayoni biroz vaqt davom etishi mumkin. Bunda kompyuterning parametrlari, internet tezligiga ham bog'liq. Bunga sabab, *docx* formatdagi matn korpusga yuklanish jarayonida, matn tokenlanadi va *json* formatga o'tkaziladi. Shu bilan birgalikda har bir so'z korpus bazasidan tekshirilib, razmetkalash jarayoni bo'ladi. Bu jarayon vaqti matnni qayta ishlashda ma'lumotlar almashinuvi katta bo'lganligi sababli uzoq bo'lishi mumkin (7-rasm). Shuning uchun matnni bo'laklarga ajratish tavsiya qilingan.



7-rasm. Matnni *avtomatik qayta ishlash*.

Matn bilan ishlash bosqichi. Soʻzlarga Grammatik maʼlumotlar berishda ikkita lugʻatdan foydalanib razmetkalaydi. Soʻzni birinchi navbatda korpus lugʻatidan tekshiradi va soʻz topilsa soʻzni *qizil rangga* oʻtkazib, razmetkalaydi. Agar soʻz korpus lugʻatidan topilmasa, uni foydalanuvchining (shaxsiy kabinet egasi) lugʻatidan izlaydi va soʻz topilsa soʻzni qizil rangga oʻtkazib, razmetkalaydi. Agar bu lugʻatdan ham soʻz topilmasa, *qora rangda* qoladi va bu soʻz foydalanuvchi tomonidan razmetkalanadi (8-rasm)[Tursunov M.S., 2022:58-63].



8-rasm. Korpusda matnning koʻrinishi.

Qizil rangda koʻrsatilgan soʻzlarni ustidan sichqoncha bilan bosilganda, soʻzning *grammatik maʼlumotlari* koʻrsatiladi (9-rasm). Agar matnning tarkibida joylashgan oʻrnida soʻzning grammatik maʼlumotlari notoʻgʻri boʻlsa *“oʻzgartirish”* tugmasi bosiladi va razmetkalash oynasiga oʻtiladi (11-rasmga qarang).

Tohir Malik
ODAMIYLIK MULKI
(Axloq kitobi)
Birinchi kitob
Toshkent «Istiqloq» 2005
Ma'ul muharrir:
Muhammad Sharif JUMAN
Taqrizchi: filologiya fanlari nomzodi Abdumurod TILAVOV
Bismiylahir Rohmanir Rohiyim.
«Darhaqiqat, Biz Odam bolalarini aziz-mukarram qildik va ularni barru- bahrda - quruqlik va dengizda (ot-ulov va kemalarga) chiqarib qo'ydik hamda ularni halol-pok narsalardan rizqu ro'z berdik va ularni O'zimiz yaratgan juda ko'p jonzotlardan afzal-ustun qilib qo'ydik.»
(Al-Isro surasidan)
GULISTON ISTA, EY KO'NGIL
Muqaddima o'rnidagi bu dastlabki bob mazkur kitobga tartib berish sabablari va maqsadlari bayonidan iboratdir.
Assalomu alaykum, azizlari
Sizlaru bizlari go'zal surat
Parvardigorning buyuk irod
salavot va salomlar yo'llayr
Yaratganning O'zi xayrli qilg
birodarlarimizga nafi tegar,
nomda e'tiborlarigizga ha
yo'lli» deb nom berdik. Bu y
Ma'lumimgizkim, o'tmishda
kuzatsak, o'tmish adabiyotimizning asosini odob masalalari band etgan. «Adabiyot» - «odoblarning to'plami» ma'nosini anglatishi ham shundandir. Badiiy asar qaysi tarzda, qaysi yo'nalishda yozilmasin, odamlarni tarbiya qilishga, ya'ni yaxshilik sari boshlab, yomonlikdan qaytarishga xizmat etishi lozim. Tarixda bunday kitoblarning hisobi yo'q. Shularning ayrimlarini o'qish jarayonida kaminada undagi hikmatlar bilan birodarlarni tanishtirish mayli

kitobga

So'z turkumi:	ot
Lemmasi:	kitob
Razmetka:	tur.ot, ya.ot, birls, o'p.k

o'zgartirish

abbimiz Alloh taolaga cheksiz hamdu sanolar aytamiz. Buyuk sizni bizga hidoyat yo'lini ko'rsatgan suyuqlik payg'ambarimiz Muhammadga «niyatimizni amalga oshirmoqqa kirishaymizkim, bu ishimizni ham monmi mashqlar qilib, o'zimizni sinab ko'rdik. Sog' odob bobida bir kitobchaga tartib berdik. Sog' uni to'ldirib, «Imonlashish umidi» degan lali bo'lib qolar, degan maqsadda bir risola yozib unga «Jinoyatning uzun isolasi nihoyasiga yetkazar, insha Alloh!

9-rasm. So'zning morfologik xususiyati.

Matnda qora rangdagi so'zlarda grammatik ma'lumotlar mavjud emas (10-rasm). Qora rangdagi so'zning ustidan sichqonchani chap tugmasi bosilganda, so'zni razmetkalash oynasi ochiladi (11-rasmga qarang).

Tohir Malik
ODAMIYLIK MULKI
(Axloq kitobi)
Birinchi kitob
Toshkent «Istiqloq» 2005
Ma'ul muharrir:
Muhammad Sharif JUMAN
Taqrizchi: filologiya fanlari nomzodi Abdumurod TILAVOV
Bismiylahir Rohmanir Rohiyim.
«Darhaqiqat, Biz Odam bolalarini aziz-mukarram qildik va ularni barru- bahrda - quruqlik va dengizda (ot-ulov va kemalarga) chiqarib qo'ydik hamda ularni halol-pok narsalardan rizqu ro'z berdik va ularni O'zimiz yaratgan juda ko'p jonzotlardan afzal-ustun qilib qo'ydik.»
(Al-Isro surasidan)
GULISTON ISTA, EY KO'NGIL
Muqaddima o'rnidagi bu dastlabki bob mazkur kitobga tartib berish sabablari va maqsadlari bayonidan iboratdir.
Assalomu alaykum, azizlari
Sizlaru bizlari go'zal suratda yaratgan va cheksiz ne'matlarga burkagan Rabbimiz Alloh taolaga cheksiz hamdu sanolar aytamiz. Buyuk Parvardigorning buyuk irodasi ila buyuk vazifalarni yelkalarida ko'targan va sizni bizga hidoyat yo'lini ko'rsatgan suyuqlik payg'ambarimiz Muhammadga salavot va salomlar yo'llaymiz. Allohning inoyati va marhamati ila ko'p yillik niyatimizni amalga oshirmoqqa kirishaymizkim, bu ishimizni ham Yaratganning O'zi xayrli qilg'ay. Shu paytga qadar badiiy ijodda yaxshimi-yomonmi mashqlar qilib, o'zimizni sinab ko'rdik. Sog' odob bobida birodarlarimizga nafi tegar, degan umid bilan «Mehmon tuyg'ular» nomida bir kitobchaga tartib berdik. Sog' uni to'ldirib, «Imonlashish umidi» degan nomda e'tiborlarigizga havola etdik. Keyin esa farzandlar tarbiyasida foydali bo'lib qolar, degan maqsadda bir risola yozib unga «Jinoyatning uzun yo'lli» deb nom berdik. Bu yo'nalishdagi urinishlarimizni endi mazkur odob risolasi nihoyasiga yetkazar, insha Alloh!

o'zgartirish

Ma'lumimgizkim, o'tmishda o'tgan ustozlarimizning aksari badiiy asarlar yozish bilan birga axloqqa doir risolalar ham bitganlar. Agar ziyraklik bilan kuzatsak, o'tmish adabiyotimizning asosini odob masalalari band etgan. «Adabiyot» - «odoblarning to'plami» ma'nosini anglatishi ham shundandir. Badiiy asar qaysi tarzda, qaysi yo'nalishda yozilmasin, odamlarni tarbiya qilishga, ya'ni yaxshilik sari boshlab, yomonlikdan qaytarishga xizmat etishi lozim. Tarixda bunday kitoblarning hisobi yo'q. Shularning ayrimlarini o'qish jarayonida kaminada undagi hikmatlar bilan birodarlarni tanishtirish mayli uyg'ondi. Rahmatli tog'am, ulug' adib Mirzakalon Ismoilov axloq masalalariga doir bir qancha risolalar yozgan edilar. Shulardan biri - «Qizlar daftariga»

10-rasm. Razmetkalanmagan so'z.

So'zni razmetkalash oynasida so'z turkumi, so'zning o'zagi, lemmasi va izohi kiritiladi. Hamma maydonlar to'ldirilgach, OK tugmasi bosiladi.

11-rasm. So'zni razmetkalash oynasi va so'z turkumini tanlash.

Tanlangan so'z turkumiga mos **grammatik teglar** hosil bo'ladi. Bunda so'zga tegishli grammatik teglar tanlanadi va **OK** tugmasi bosiladi. Undan so'ng **razmetka maydoni** to'ldirilgan holatda bo'ladi va bu maydonga o'zgartirish kiritish mumkin emas. (12-rasm).

12-rasm. So'z teglash.

So'z razmetkalangach **ko'k rangga** o'zgaradi. Bu so'zga foydalanuvchi tomonidan grammatik ma'lumotlar berilganini bildiradi. Bunda so'z foydalanuvchi lug'atiga qo'shiladi (13-rasm).

Tohir Malik
ODAMIYLIK MULKI
(Axloq kitobi)
Birinchi kitob
Toshkent «Istiqloq» 2005
Ma'ul muharrir:
Muhammad Sharif JUMAN
Taqrizchi: filologiya fanlari nomzodi Abdumurod TILAVOV
Bismiylahir Rohmanir Rohim.
«Darhaqiqat, Biz Odam bolalarini aziz-mukarram qildik va ularni barru- bahrda - quruqlik va dengizda (ot-ulov va kemalarga) chiqarib qo'ydik hamda ularni halol-pok narsalardan rizqu ro'z berdik va ularni O'zimiz yaratgan juda ko'p jonzorlardan afzal-ustun qilib qo'ydik.»
(Al-Isro surasidan)
GULISTON ISTA, EY KO'NGIL
Muqaddima o'rnidagi bu dastlabki bob mazkur kitobga tartib berish sabablari va maqsadlari bayonidan iboratdir.
Assalomu alaykum, azizlarim.
Sizlaru bizlarni go'zal suratda yaratgan va cheksiz ne'matlarga burkagan Rabbimiz Alloh taolaga cheksiz hamdu sanolar aytamiz. Buyuk Parvardigorning buyuk irodasi ila buyuk vazifalarni yelkalarida ko'targan va sizni bizga hidoyat yo'lini ko'rsatgan suyuqlik payg'ambarimiz Muhammadga salavat va salomlar yo'llaymiz. Allohning inoyati va marhamati ila ko'p yillik niyatimizni amalga oshirmoqqa kirishyapmizkim, bu ishimizni ham Yaratganning O'zi xayrli qilg'ay. Shu paytga qadar badiiy ijodda yaxshimi-yomonni mashqlar qilib, o'zimizni sinab ko'rdik. So'ng odob bobida birodarlarmizga nafi tegar, degan umid bilan «Mehmon tuyg'ular» nomida bir kitobchaga tartib berdik. So'ng uni to'ldirib, «Imonlashish umidi» degan nomda e'tiborlaringizga havola etdik. Keyin esa farzandlar tarbiyasida foydali bo'lib qolar, degan maqsadda bir risola yozib unga «Jinoyatning uzun yo'li» deb nom berdik. Bu yo'nalishdagi urinishlarimizni endi mazkur odob risolasi nihoyasiga yetkazar, insha Alloh!
Ma'lumimgizkim, o'tmishda o'tgan ustozlarimizning aksari badiiy asarlar yozish bilan birga axloqqa doir risolalar ham bitganlar. Agar ziyraklik bilan kuzatsak, o'tmish adabiyotimizning asosini odob masalalari band etgan. «Adabiyot» - «odoblarning to'plami» ma'nosini anglatishi ham shundandir. Badiiy asar qaysi tarzda, qaysi yo'nalishda yozilmasin, odamlarni tarbiya qilishga, ya'ni yaxshilik sari boshlab, yomonlikdan qaytarishga xizmat etishi lozim. Tarixda bunday kitoblarning hisobi yo'q. Shularning ayrimlarini o'qish jarayonida kaminada undagi hikmatlar bilan birodarlarni tanishtirish mayli uyg'ondi. Rahmatli tog'am, ulug' adib Mirzakalon ismoliy axloq masalalariga doir bir qancha risolalar yozgan edilar. Shulardan biri - «Qizlar daftariga»

13-rasm. Tadqiqotchi tomonidan razmetkalangan so'z holati.

Matnlarni korpusga kiritish uchun, uni raqamlashtirish va korpus talabiga ko'ra formatlashtirish lozim bo'ladi. Raqamlashtirish jarayonida ko'proq matnni lingvistik xatolarini tuzatishga e'tibor berish lozim. Bu esa foydalanuvchiga lingvistik ma'lumotlarni yuqori aniqlikda olish imkonini beradi. Umuman olganda, matnni korpusga kiritishdan avval matnning imloviy va texnik xatolarini asl matn bilan tekshirib to'g'rilanadi. Bu korpusning lug'atida va korpusga kiritilgan matnning sifatini yaxshilaydi hamda korpusda keraksiz va ortiqcha ma'lumotlar yig'ilishining oldi olinadi.

Foydalanilgan adabiyotlar:

1. Colm Smyth, "An introduction to corpus linguistics", Bulletin of Tokyo Denki University, Arts and Sciences No.14, 2016, 105-109 p.
2. McEnery T., Hardie A., "Corpus linguistics: method, theory and practice", Book, Cambridge University Press, 2012, 294 p.
3. Qarshiyev A.B., Karimov S.A., Tursunov M.S., "O'zbek tili milliy korpusi uchun matnlarni formatlash", Raqamli texnologiyalarning nazariy va amaliy masalalari xalqaro jurnali, TATU Samarqand filiali, Samarqand, №1(1)2022, 58-63 s.
4. Tursunov M.S., "Description of the management system programs of the national corpus of the uzbek language", International journal of engineering mathematics, <https://iejemta.com/index.php/em/article/view/9/9>, volume 4 ISSUE 1, 2022, 32-42 p.
5. Бочаров В.В., Грановский Д.В., "Программное обеспечение для коллективной работы над морфологической разметкой корпуса", Труды международной конференции «Корпусная лингвистика — 2011». СПб.: Изд-во С.-Петерб. ун-та, 2011. С. 104–109.

СОЗДАНИЕ И ЗНАЧЕНИЕ ЯЗЫКА КОРПУС В УЗБЕКИСТАНЕ

Тоирова Гули Ибрагимовна

доктор филологических наук, профессор,

Бухарский государственный университет

ORCID 0000-0003-1794-6575

E-mail: tugulijon@mail.ru

Аннотация: В статье рассматривается трансформация языка в язык Интернета, компьютерные технологии, математическая лингвистика, ее продолжение и становление и развитие компьютерной лингвистики, в частности вопрос моделирования естественных языков для искусственного интеллекта. Определены требования к кодированию важной текстовой информации. Состояние анализирует лингвистический модуль и алгоритм и его виды из независимых компонентов лингвистического программного кода. Научно обоснована необходимость алгоритмов фонологических, морфологических и орфографических правил формирования лексико-грамматического кода. Подчеркивается важность таких лингвистических модулей, как фонология, морфология и орфография, в формировании лингвистической базы национального корпуса узбекского языка. В статье рассматривается основное назначение корпуса как сложного лингвистического источника, а также тот факт, что он в основном содержит два вида информации и ее типы.

Ключевые слова: корпус, искусственный интеллект, лексическая информация, морфологический признак, слово.

Annotation. The article discusses the transformation of language into the language of the Internet, computer technology, mathematical linguistics, its continuation and the formation and development of computational linguistics, in particular, the issue of modeling natural languages for artificial intelligence. The requirements for coding important textual information are determined. The state analyzes the linguistic module and the algorithm and its types from the independent components of the linguistic program code. The need for algorithms for phonological, morphological and spelling rules for the formation of a lexico-grammatical code is scientifically substantiated. The importance of such linguistic modules as phonology, morphology and spelling in the formation of the linguistic base of the national corpus of the Uzbek language is emphasized. The article discusses the main purpose of the corpus as a complex linguistic source, as well as the fact that it mainly contains two types of information and its types.

Keywords: corpus, artificial intelligence, lexical information, morphological feature, word.

Благодаря современным информационным технологиям искусственный интеллект дал широкий спектр преимуществ в использовании языка. Он способен делать множество вещей, на которые способен человеческий интеллект. Электронные источники, созданные искусственным интеллектом, предназначены для обеспечения безопасности людей и снижения их веса. Среди наиболее актуальных проблем - перевод узбекского языка на Интернет и электронный язык, а также расширение электронных ресурсов на национальных языках (корпус узбекских языков, электронные словари, содержание веб-сайтов).

В мировой лингвистике создание языковых корпусов в Интернете является основным средством сохранения того или иного языка ко второму десятилетию двадцать первого века, расширения сферы его исследования и демонстрации языковых навыков. В частности, компьютерная технология, которая является великим изобретением двадцатого века, открывает дверь в широкий спектр возможностей для лингвистики, а также других областей и накладывает огромные задачи на компьютерный язык, появление компьютерной лингвистики имеет решающее значение для успех естественных языков.

Корпус является предметом корпусной лингвистики. Этот термин по-разному определяется в научной литературе. Например, в английском языке он используется с такими терминами, как лингвистический корпус или текстовый корпус. Признание научных исследований А.Н. Хомский, Г.Н.Луч, Ч.Ф.Меер, Дж.Синклер, М.З.Курди в решении таких проблем, как создание национального корпуса определенного языка, его аналитическая технология, развитие области корпусной лингвистики должны [Мохамед Закария Курди, 2016 г. ; Тоирова, 2020, с.57; Чарлез, 2004, с.7; Шомску, 1962].

Джон Синклер определяет термин «корпус» следующим образом: «Корпус состоит из фрагментов текстов в электронной форме, отобранных по видимым критериям для изучения языка или языкового разнообразия, которые должны быть представлены в качестве источника информации» [Sinclair, 2004].

Большой массив массивных текстов русской корпусной лингвистики, принципы формирования корпусов, лингвистическая база данных В.Г. Бритвин, В.П. Захаров, И.А. Мельчук, А.Б. Кутузов, Р.Г. Котов, Л.И. Беляева, Отраженные в целевых исследованиях Е.В.Недошивиной, В.В.Рыкова, В.Плунгяна [Бритвин, 1983; Блумфилд, 1968; Беляева, Чижаковский, 1983; Захаров, 2011; Недошивина, 2006; Рыков, 2005; Плунгян, 2005; Кутузов, 2017; Котов, 1977].

Русский ученый В.П. Захаров объясняет термин «корпус» следующим образом: «корпус — совокупность лингвистических данных единиц языка, составленных на основе устных и письменных текстов» [Захаров, 2011].

Х.Исхакова, С.Мухаммедов, С.Риза о лингвостатистическом анализе текста в узбекском языкознании, лексикографической обработке, лингвистическом обеспечении программы автоматического редактирования, лингвистических модулях программы редактирования и анализа, синонимической лексике национального корпуса, лингвистические основы авторского корпуса. С.Мухаммедова, Б.Менглиев, Д.Уринбаева, А.Пулатов, У.Дысимова, Г.Валиева, Г.Джуманазарова, Н.Абдурахмонова, Ш.Хамроева, М.Абжалова, А.Эшмоминов, О.Холиёров, Р. Работа Каримова заслуживает внимания. Наши ученые, такие как С.Каримов, С.Мухаммедова, Ш.Хамроева, проводили исследования по специальности 10.00.01 корпусной лингвистики.

Языковые корпуса можно разделить на разные формы по структуре, назначению, устойчивости, вариативности. Например, В.П. Захаров перечисляет следующие формы: «по форме хранения данных: звуковая, письменная, смешанная; по языку текста: одноязычные и многоязычные; по жанру: литературный, диалектный, устный, публицистический, смешанный; по подъезду к зданию: свободный, хозяйственные постройки, закрытый; по назначению: исследовательская, иллюстративная; по изменчивости: динамичные и стабильные; маркированные и не маркированные по признаку наличия дополнительной информации (аннотированные)» [Захаров, 2011].

В.В. Рыков же акцентирует внимание на следующих аспектах классификации корпусных типов: «По уровню и структуре данных, по хронологическому признаку (позиции) языка, по языку употребления, по до цели использования» [Рыков, 2005].

В своем запрете Ш. Хамроева делит корпус на следующие типы: «По определенному периоду языка или определенному типу его возникновения (жанр, стиль, социальная или возрастная группа, язык писателя или ученого); по типу языкового знака; по типу речи: письменная, устная, смешанная; они выглядят как мультимодальный корпус, корпус специальных текстов» [Хамроева, 2018].

«Специализированный корпус: группа текстов определенного типа: газетный текст, научные статьи; общий корпус; сравнительный корпус; параллельный корпус; учебный корпус; дидактический корпус», — говорит У.Холиёров [Холиёров, 2021].

Каждый академик определял лингвистические корпуса со своей точки зрения и классифицировал их по-разному. Какие черты узбекского лингвистического корпуса нашли отражение в нем и какие корпуса сейчас создаются?

Создание национального корпуса узбекского языка является относительно новым направлением как в узбекской лингвистике, так и в современных информационных технологиях. Языковой корпус является основным источником и мощным информационным ресурсом для составления больших словарей. Языковой корпус позволяет быстро создавать и обрабатывать словари с помощью компьютера. Важность корпуса в области лексикографии заключается в том, что ни один инструмент не может сравниться с корпусом в определении периода и частоты использования слова. В ближайшем будущем потребность в словаре сегодня для студента, изучающего язык, или исследователя, изучающего какой-либо его аспект, несомненно, переместится на корпус.

Лингвисты Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои сейчас работают над научным и практическим проектом под названием «Образовательный корпус». Создание образовательного корпуса по узбекскому языку направлено на поэтапное формирование данных на основе зарубежного опыта и включает в себя электронный учебник, содержащий современную лексику узбекского литературного языка, многоязычных носителей и непереведенных лексических единиц узбекского языка, а также комплект мультимедийных продуктов, включающий аудио- и видеоматериалы, а также мобильное приложение, направленные на формирование навыков правильного произношения на узбекском языке.

Участие в международной научно-практической конференции «Теоретические и практические вопросы создания узбекского национально-образовательного корпуса» в мае 2021 года, инициированной учеными Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои, с практическими предложениями не только отечественными, но и тоже международный сделал. По итогам конференции были определены конкретные задачи по созданию качественного национального и образовательного корпуса с привлечением международного опыта.

Ученые Самаркандского государственного университета также работают над проектом под названием «Проектирование и разработка национального корпуса узбекского языка», а ученые Бухарского государственного университета проводят исследования по теоретическим и практическим вопросам создания национального корпуса узбекского языка. Результаты исследования имеют большое значение в повышении международного

статуса узбекского языка, поднятии его до уровня мирового языка общения, изучении и преподавании узбекского языка за рубежом, расширении возможностей нашего национального языка.

Создание национального корпуса узбекского языка позволит «оцифровать» узбекский язык и превратить его в интернет-язык.

Короче говоря, корпусная лингвистика — это самая передовая ветвь лингвистики, а корпус — необходимый инструмент для лингвистов; устные, письменные памятники являются источником информации, отражающей национально-культурное наследие. Корпус представляет собой совокупность текстов, подлежащих поисковой программе, а четко определенный корпус служит устойчивой лингвистической базой в обеспечении эффективности лингвистического исследования. Являясь продуктом искусственного интеллекта, лингвистический корпус включает в себя электронный словарь, портал переводов, терминологическую базу, виртуальную (электронную) библиотеку, электронное правительство, электронные издательства, электронные учебники и учебные пособия. Общий вид узбекского национального корпуса разделен на несколько окон и правую и левую колонки. В нем будут следующие окна: «Лексический поиск», «Морфологический поиск», «Синтаксический поиск». Слова и фразы из него автоматически анализируются за считанные секунды. Будет проведен пересмотр теоретических основ морфологической и синтаксической разметки на основе академической грамматики, практическая работа, связанная с сокращением системы семантических тегов разметки.

Использованная литература:

1. Eshmuminov, A., (2019). Synonymous database of the Uzbek language national corpus. Dissertation of PhD in Philology, Tashkent.
2. Fries, Ch.C., (1969). The structure of English. An introduction to the construction of English sentences, London.
3. Hamroeva, Sh., (2018). Linguistic bases of creation of the author's corpus of the Uzbek language: Author's Abstract of the Dissertation of PhD in Philology, Tashkent.
4. Karimov, R., (2021). Linguistics and programming issues of creating a parallel corpus of Uzbek and English, Author's Abstract of dissertation of PhD, Bukhoro, 151 p.
5. Kholiyorov, O., (2021). Linguistic bases of formation of educational corpus of Uzbek language. Author's Abstract of the Dissertation of PhD in Philology, Termiz.
6. Kotov, R.G., (1977). Linguistic aspects of automated control systems. Moscow, Nauka.
7. Kutuzov, A.B., (2017). Corpus linguistics. Retrieved from: <http://lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf> ;
8. Leech, G., (1991). The State of Art in Corpus Linguistics, English Corpus Linguistics, London.
9. Melchuk, I.A., (1985). Word order in the automatic synthesis of the Russian word (preliminary messages), Scientific and technical information, 12:12-36.
10. Mohamed Zakaria Kurdi, (2016). Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax, Great Britain, USA: Wiley-ISTE, 300 p.
11. Pulatov, A. Q., (2011). Computer Linguistics. Tashkent, Akademnashr, 520 p.

12. Rykov, V.V., (2005). A course of lectures on corpus linguistics. URL: <http://rykov-cl.narod.ru/c.html>.
13. Toirova G., Yuldasheva M., Elibaeva I. Importance of Interface in Creating Corpus. // International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S10, September 2019. –P.352-355.
14. Toirova, G., (2019). The Role of Setting in Linguistic Modeling. International Multilingual Journal of Science and Technology, 4(9):722-723, available at: <http://imjst.org/index.php/vol-4-issue-9-september-2019/>.
15. Toirova, G., (2020). About the technological process of creating a national corpus. Foreign languages in Uzbekistan, 2(31):57-64, available at: <https://journal.fledu.uz/uz/2-31-2020>.
16. Toirova, G., (2020). The importance of the interface in the creation of the corpus. Internauka, 7, DOI: <https://doi.org/10.25313/2520-2057-2020-7-5944>.
17. Toirova G. The Role Of Setting In Linguistic Modeling. //International Multilingual Journal of Science and Technology. ISSN: 2528-9810 Vol. 4 Issue 9, September – 2019, – P.722-723
18. Тоирова Г. Ўзбек тили миллий корпусини яратишда лингвистик разметканинг ўрни. // Бухоро Давлат университети илмий ахбороти. – Бухоро, 2020. – №1 (77). – Б.125-132.
19. Тоирова Г. Корпус лингвистикасининг атамалар луғати. Изоҳли луғат. Германия: «GlebeEdit» номли халқаро нашриёт. 2020, –68 б.
20. Zakharov, V.P., (2011). Corpus linguistics: a textbook for students of humanitarian universities, Irkutsk, 161 p.
21. N. Z. Abdurakhmonova, A. S. Ismailov and D. Mengliev, «Developing NLP Tool for Linguistic Analysis of Turkic Languages,» 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation, 2022, pp. 1790-1793, doi: 10.1109/SIBIRCON56155.2022.10017049.
22. 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.
23. 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
24. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. In Proceedings of the International Conference on Language Technologies for All (LT4All) 2019.

OG‘ZAKI NUTQ KOMPYUTER KORPUSLARI

Abduraxmonova Nilufar Zaynobiddin qizi

O‘zbekiston Milliy universiteti professori,
filologiya fanlari doktori

E-mail: n.abduraxmonova@nuu.uz

Rajabova K., Xasanov S.,

Jumanazarova D., O‘roqova D., Yo‘ldosheva R.

O‘zbekiston Milliy universiteti
Kompyuter lingvistikasi mutaxassisligi magistrarlari

Annotatsiya: Ushbu maqolada audio korpus yaratish bo‘yicha dunyo tajribasida sinalgan namunalar tahlilga tortilgan. Audiokorpus yaratish doirasida korpus lingvistikasi oldida turgan asosiy vazifalar, lingvistik masalalar ushbu maqolada o‘z aksini topgan. Tabiiy tilni qayta ishlash va kompyuter lingvistikasida nutqiy sintezatorlarni yaratish, qolaversa, kompyuter lingvodidaktikasini rivojlantirish ovozli korpuslarning o‘rni haqidagi mulohazalar bildirilgan.

Kalit so‘zlar: korpus lingvistikasi, audiokorpus, lingvodidaktika, metama’lumot.

Annotation: Examples in the world experience of creating an audio corpus are analyzed in this article. The main tasks and linguistic issues facing corpus linguistics within the framework of creating an audio corpus are represented in this article. The creation of speech synthesizers in natural language processing and computer linguistics, as well as the role of audio corpora in the development of computer linguistics are discussed.

Keywords: linguistic corpus, audio corpus, linguistics, metamaterial.

Аннотация. В статье анализируются примеры из мирового опыта создания аудиокорпусов. В статье представлены основные задачи и лингвистические проблемы, стоящие перед корпусной лингвистикой в рамках создания аудиокорпуса. Обсуждается создание синтезаторов речи в обработке естественного языка и компьютерной лингвистике, а также роль аудиокорпусов в развитии компьютерной лингводидактики.

Ключевые слова: корпусная лингвистика, аудиокорпус, лингводидактика, метаданные.

Og‘zaki nutq korpusi beixtiyoriy yozib olingan yoki rejalashtirilgan nutqning transkripsiyalarini o‘z ichiga oladi, masalan, efirga uzatilgan yangiliklar yoki e‘lon qilingan hikoyalar va dialoglar. Audio matnlar ko‘pincha yozma matnlar bilan birgalikda qo‘llanadi. Ular fonologiya, so‘zlashuv tahlili va dialektologiya kabi turli xil tilshunoslik tadqiqotlari uchun bebaho manba hisoblanadi. Bunday korpuslar uchun tanlangan materiallar sinchkovlik bilan tanlanadi va sotsiodemografik metama’lumotlar bilan tavsiflanadi.

CLARIN infratuzilmasida 133 ta og‘zaki nutq korpusi mavjud bo‘lib, ulardan 122 tasi og‘zaki yoki spontan nutqning transkripsiyalarini va tegishli yozuvlarni, 11 tasi esa faqat transkripsiyalarni o‘z ichiga oladi. Korpuslarning aksariyat qismi bir tili bo‘lib, quyidagi 15 ta

tilni o'z ichiga oladi: arab, chex, golland, eston, fin, fransuz, nemis, venger, italyan, nepal, norveg, polyak, skoti saami, sloven, ispan va shved. Aksariyat hollarda korpusni to'g'ridan-to'g'ri milliy resurslardan yuklab olish yoki ulardan foydalanish oson onlayn qidiruv muhitlari orqali foydalanish mumkin. Ular, shuningdek, fonemik va prosodik annotatsiya kabi nutq korpusiga xos belgilarga ko'ra annotatsiyalangan bo'ladi [<https://www.clarin.eu/resource-families/spoken-corpora>].

Yozma matn korpusidagi matnlar hajmi og'zaki nutq korpusiga qaraganda ancha ko'p. Yozma matnga nisbatan og'zaki matnga xos bo'lgan belgilarga alohida e'tibor qaratish juda muhim sanaladi. Bunga imo-ishoralar, ohang, diskurs strukturasi kabilar yordamida og'zaki matnni tahlil qilishda foydalaniladi [Robbie Lovei: 2017].

Britaniya Milliy korpusining og'zaki matn korpusi 1991 yildan 1994-yilgacha bo'lgan davrda yaratilgan. U ikki qismdan iborat: 1- qism 40% demografik namunlardan olingan, 2- qism esa 60%i kontekstga asoslangan. Shuningdek, ushbu korpusda norasmiy, doimiy so'zlashuv uslubiga xos og'zaki matnlar kiritilgan. Yozib olingan audio hujjatlar ko'ngillilar tomonidan yozib olingan bo'lib, har ular yosh, jins, ijtimoiy guruh va geografik region jihatdan tasniflangan. 1000dan ortiq kishilar tomonidan yozib olingan audio fayllar 4.2 mln. so'z transkripsiya qilingan. Britaniya milliy korpusining og'zaki nutq korpusi 1993-yildagi London maktablaridagi 13-17 yoshli o'smirlar o'rtasidagi og'zaki suhbatlari yarim mln so'z shakllaridan tashkil topgan London o'smirlar tilining Bergen korpusi (COLT) bilan ham hamkorlik qiladi [Stenström, A: 2002]. Ayrim korpus dizayni va metama'lumotni berishdagi ayrim kamchiliklari bilan bo'lsa-da, uzoq yillar lingvistik tadqiqotlar uchun yuqori natijali resurs bo'lib xizmat qiladi.

Shuningdek, ingliz tiliga doir boshqa og'zaki nutq korpuslari ham mavjud. Masalan, Britaniya orolidagi so'zlovchilar nutqiga tegishli SCOTS shodland ingliz tilidagi bit mln.so'zdan iborat (2000)[Douglas, F.: 2003]. Yana bir korpus turi bu BASE (the British Academic Spoken English Corpus) Britaniya akademik og'zaki ingliz tili korpusi universitet ma'ruzalari va seminarlaridan tashkil topgan [Thompson, P.: 2001].

Og'zaki nutq korpusiga oid ma'lumotlar muayyan tadqiqot maqsadi hamda muayyan sohadan kelib chiqiladi. Shu nuqtai nazaridan BNC, ya'ni Britaniya milliy korpusi demografik kategoriyasi ikki turga ajratilgan: tanlash tamoyili (selection criteria) va diskriptiv (tavsifiy) tamoyil (descriptive criteria) [Robbie Lovei.:2017]. Tanlash tamoyiliga so'zlovchilarning jins, yosh, sotsio-iqtisodiy mavqei va hududiy xoslanishi kiradi. Diskriptiv tamoyilga yozib olingan nutq turi va muayyan sohani qamrab olgan ma'lumot tanlov jarayonida nazorat qilinmaydi. BNCning tanlov tamoyiliga ko'ra quyidagicha tasniflangan¹:

Tanlov tamoyili	Kategoriyasi	% "so'z ulushi"
Jins	Erkak	41.14
	Ayol	58.47
	noma'lum	0.38
Yosh	0-14	6.30
	15-24	15.71
	25-34	20.16

¹ O'sha joyda. – P. 326.

Ijtimoiy-iqtisodiy status

35-44	19.96
45-59	22.76
60+	15.08
AB	32.41
C1	26.08
C2	25.69
DE	14.91
Noma'lum	0.88

BNCning 2014yilda yig'ilgan og'zaki nutq ma'lumotlari dialektlar kesimida tasniflangan:

Global	Davlat	Ichki region	Region
Buyuk Britaniya	Ingliz	Shimol	Shimoli sharq
			Yorkshar va Hambersayd (Yorkshire, Humberside)
			Shimoli g'arb
			Merseyside
		Midlands	Sharqiy Midlands
			G'arbiy Midlands
		Janub	Sharqiy
			Janubi g'arb
			Janubi sharq
	Shotlandiya	Shotland	Shotlandiya
Buyuk Britaniyalik bo'lmagan	Uels	Uels	Uels
	Shimoliy Irlandiya	Shimoliy Irlandiya	Shimoliy Irlandiya
	Irlandiya	Irlandiya	Irlandiya
	Buyuk Britaniyalik bo'lmagan	Buyuk Britaniyalik bo'lmagan	Buyuk Britaniyalik bo'lmagan

og'zaki korpus bugungi kunda tilning nutqdagi voqelanishi sifatida muhim resurs vazifasini bajarmoqda.

Og'zaki matn korpuslarini yaratish umumiy tamoyil asosida yaratilishi kerak. Sinkler quyidagi yo'riqnomani keltiradi [Sinclair, J.:2005]:

1. Korpus kontenti tilda nima mavjud bo'lishi emas, balki jamiyatda uning kommunikativ funksiyasiga qarab tanlanishi kerak;
2. Korpus tuzuvchilari tanlangan tilning reprezentativlik bo'lish ehtimolligini ta'minlashi kerak;
3. Korpusning komponentlari mustaqil shaklda loyihalashtirilishi lozim;
4. Korpus strukturasi kam sonli bo'lib, tekshirish uchun alohida yaratilishi kerak;
5. Matn haqidagi ma'lumotlar alohida saqlanishi va zaruriyat tug'ilganda ilovalarda oddiy matndan alohida saqlanishi kerak;

6. Korpus uchun til namunalari to'liq holda hujjatlardan yoki tugallangan nutqiy aktlardan iborat bo'lishi lozim. Namunalar hajm jihatdan bir-biridan farqlanishi kerak;
7. Korpusning dizayni va tuzilishi uning tarkibi va argumentlari haqidagi ma'lumotlar to'liq hujjatlashtirilishi lozim;
8. Korpus tuzuvchilar korpus representativli va muvozanatni hisobga olishlari lozim. Agar maqsad aniq bo'lmasa, ularning korpus komponentlari va dizayni haqida yo'riqnomasi bo'lishi kerak;
9. Korpusdagi biror masalani hal qilish ichki talab bo'yicha emas, tashqi talab asosida bajarilishi kerak;
10. Korpusdagi komponentlarning bir xilligini ta'minlashda ishonchsiz matnlardan saqlanishi kerak.

Yuqorida keltirilgan korpus yaratish bo'yicha umumiy talablar og'zaki matn korpusi uchun ham muayyan darajada amal qiladi. Biroq ayrim jihatlari bilan bir-biridan farq qiladi:

- Ovozni yozish
- Transkripsiya qilish, kodlash va annotatsiyalash
- Boshqaruv va tahlil

Korpusni loyihalash qismi muhim sanalib, korpusning boshqa qismlariga ham ta'sir qiladi.

Ma'lumotlarni yozish rejlasmatirilgan tahlil turi bo'yicha aniqlanadi, so'ng qolgan bosqichlar amalga oshiriladi. Korpusdagi diskursning to'liq manzarasini yaratish mumkin bo'lmaganligi uchun yozib olinayotgan diskurs ma'lumotlarini loyihalash juda ham muhim sanaladi. Demak, korpus yaratuvchilari korpusning representativligi, bir maqsadlilik va muvozanatiga qat'iy rioya qilishlari lozim.

Korpusga ovoqli matnlarni ma'lumotlar bazasiga yig'ish, saralash va kiritish lingvistik tadqiqotlar uchun yo'naltirilganligiga ko'ra dolzarb sanaladi.

Olingan hujjatlardagi ishtirokchilar makon va umumiy olinayotgan vaziyatdagi ma'lumotlar, hattoki, ovoz yozish uskunasiining texnik va fizik xususiyatlari ham qayd etiladi. Jonli ovozni yozib olishda uchraydigan biror ma'lumotning yo'qolib qolishi yoki tushib qolgandan keyingi ma'lumotda uni qayta tuzatish oson bo'lmaganligi uchun ham yuqorida keltirilganlarning barchasiga amal qilinishi lozim.

Tadqiqotchilarning maqsadlaridan kelib chiqib yig'ilgan barcha ma'lumotlar haqida invariant qoidalar to'plamini berishning imkoni yo'q [Cameron D.:2001]. Ovozli matn soatlar miqdoriga qarab, so'zlar soni ham arxivlanishi lozim. Masalan, CONCODE loyihasi yozib olingan odatiy bir soatlik muloqot transkripsiya qilingan 10 ming so'zdan tashkil topgan. Bir soatda yozib olingan ovoqli matnda qatnashuvchi so'zlarning miqdori turli faktorlarga, xususan, diskurs konteksti va ishtirokchining nutq tezligiga bog'liq bo'ladi [Routledge handbook of corpus linguistics.:2010]. Og'zaki matn korpusining hajmi qancha bo'lishi bu tadqiqotda ko'zda tutilgan umumiy maqsaddan kelib chiqiladi.

Korpusdagi og'zaki matn yaratish uchun qatnashayotgan ishtirokchilardan tashqari, o'sha joy va vaziyat haqidagi ma'lumotlar ham zarur bo'ladi. Burnard korpus uchun quyidagi metama'lumotlarni keltirib o'tadi [Burnard L.: 30-46]:

- Tahririy metama'lumot – asliyatdagi manba bilan korpus komponentlari o'rtasidagi munosabatlar haqida ma'lumot beradi;

- Analitik metama'lumot – korpus komponentlari qaysi usulda interpretatsiya qilinishi va tahlil qilinishi bo'yicha ma'lumot;
- Deskriptiv metama'lumot – korpus komponentlarining ichki va tashqi xususiyatlari olingan tasnifiy ma'lumotlarning berilishi;
- Administrativ metama'lumot – korpusning o'zi haqidagi sarlavhasi, foydalanish imkoniyati, qayta tahrirlash statusi haqidagi ma'lumotlar keltiriladi.

Audiomatnlarni transkripsiya qilish murakkab va ko'p mehnat talab qiluvchi jarayon sanaladi. Matnlarni transkripsiya qilishda foydalanuvchilar uchun qayta ishlashga moslashtirish maqsadida qat'iy yo'riqnomaga amal qilinishi lozim. COBUILD loyihasi uchun turli transkripsiyalarni o'z ichiga olgan Network of European reference Corpora (NERC)dan foydalanilgan [Burnard L. : 30-46]. Ushbu transkripsiya tizimi nutq ohangi haqidagi ma'lumotni o'z ichiga oluvchi orfografik ifodadan tavsiflangan transkripsiya bo'lgan to'rt qismdan iborat bo'ladi. Boshqa transkripsiya qilish yo'riqnomasi Text Encoding Initiative (TEI) asosida amalga oshiriladi. Masalan, Britaniya milliy korpusi shu yo'riqnoma asosida transkripsiya qilingan. Ushbu yo'riqnoma og'zaki nutq aktining struktural, kontekstual, prosodik, temporal annotatsiyadan iborat bo'lib, muayyan tadqiqot maqsadi uchun moslashtirilgan. Fonemalarni transkripsiya qilishda Xalqaro fonetik alifbosidan foydalanilgan. Multi-modal korpusidagi (NMMC-multi-modal corpus) audio va video ovozlari kelishilgan formatda transkripsiya qilingan.

Ovozli ma'lumotlarni kodlash deganda, berilganlarni raqamli ma'lumotga aylantirish tushuniladi. Kodlash til resurslarini annotatsiyalash jarayoni bo'lib, dasturiy ta'minot orqali amalga oshiriladi. Bu kabi annotatsiyalash avtomatik usulda amalga oshiriladi. Dastlabki korpuslarni annotatsiyalash SGML (Standard Generalized Mark-up Language) 1980-yillarda yaratilgan. SGML transkripsiya va annotatsiyani olish uchun standartlarni berishda nutqiy aktlar o'rtasidagi to'xtamlar, paragraf chegaralari bo'yicha belgilar qo'yilgan.

Nutqiy aktdagi to'xtamlar ham "qisqa" va "uzun" bo'lishi mumkin [Johansson, S.:1995]. Ular turlicha teglangan:

	Qisqa to'xtam	Uzun to'xtam	Vaqt to'xtam
Du Bois (1990)	..	...	...(1.5)
MacWhinney (1991)	#	##, ###, # uzun	#1_5
Rosta (1990)	<,>	<,,>	
Svartik va Quirk (1980)	.	—	—..

Foydalanilgan adabiyotlar:

1. <https://www.clarin.eu/resource-families/spoken-corpora>
2. Robbie Lovei, Claire Dembryii, Andrew Hardiei, Vaclav Brezinai, Tony McEneryi The Spoken BNC2014 Designing and building a spoken corpus of everyday conversations // International Journal of Corpus Linguistics 22:3
3. (2017), 319–344. doi 10.1075/ijcl.22.3.02lov issn 1384-6655 / e-issn 1569-9811 © John Benjamins Publishing Company This is an open access article under a OA CC BY license
4. Stenström, A. -B., Andersen, G., & Hasund, I. K. (2002). Trends in Teenage Talk: Corpus Compilation,
5. Analysis and Findings. Amsterdam/Philadelphia: John Benjamins. doi:10.1075/scl.8

6. Douglas, F. (2003). The Scottish Corpus of Texts and Speech: Problems of corpus design. *Literary and Linguistic Computing*, 18(1), 23–37. doi:10.1093/lc/18.1.23
7. Thompson, P., & Nesi, H. (2001). The British Academic Spoken English (BASE) Corpus Project. *Language Teaching Research*, 5(3), 263–264.
8. Robbie Lovei, Claire Dembryii, Andrew Hardiei, Vaclav Brezinai, Tony McEneryi The Spoken BNC2014 Designing and building a spoken corpus of everyday conversations // *International Journal of Corpus Linguistics* 22:3 (2017), 319–344. doi 10.1075/ijcl.22.3.02lov issn 1384-6655 / e-issn 1569-9811 © John Benjamins Publishing Company This is an open access article under a OA CC BY license
9. Sinclair, J., 2005. Corpus and text: Basic principles. In: M. Wynne (ed.), *Developing Linguistic Corpora*:
10. *A Guide to Good Practice*, Oxford: Oxbow Books, 1-16, available online at
11. <http://ota.ahds.ac.uk/documents/creating/dlc/chapter1.htm> [date accessed 25-06-2012]. Cameron D. *Working with spoken Discourse*. London: Sage. 2001.
12. *Routledge handbook of corpus linguistics*, USA. 2010 / Svenja Adolphs, Dawn Knight Building a spoken corpus: What are the basics- P. 41.
13. Burnard L. *Developing linguistic Corpora: Metadata for corpus work*, in M. Wynne (ed.) *Developing linguistics corpora: A guide to good practice*. Oxford: Oxbow books, pp. 30-46
14. Johansson, S. 1995. The approach of the Text Encoding Initiative to the encoding of spoken discourse. In *Spoken English on Computer*, eds. G. Leech, G. Myers and J. Thomas, 82-98. Harlow: Longman

O'ZBEK TILI KORPUSI DASTURIY TA'MINOTINI YARATISH**Sharipov Maqsud Siddiqovich**

Doctor of Philosophy, Docent

Urganch State University

E-mail: maqsbek72@gmail.com**Kurbanova Lola Ulug'bek qizi**

Teacher, Urganch State University

E-mail: lolakompyuterlingvistikasi@gmail.com**Qurbanova Ro'zikajon Ulug'bek qizi**

Graduate master student, Urganch State University

E-mail: Ruzikajonqurbanova0809@gmail.com

Annotatsiya: Til korpusi - lingvistik tahlil qilish uchun tuzilgan va izohlangan yozma yoki og'zaki matnlar to'plamidir. Ushbu korpuslar tilshunoslik, tabiiy tilni qayta ishlash (NLP) va kompyuter lingvistikasi kabi sohalaridagi tadqiqotchilar uchun foydali manba bo'lib xizmat qiladi. Bu til uchun raqamli resurslar va lingvistik vositalarning cheklanganligini hisobga olsak, o'zbek tili uchun korpus yaratish juda qiyin vazifadir. Ushbu maqolada O'zbek tilining milliy korpusi dasturiy ta'minotini yaratishning nazariy va amaliy asoslari o'rganilgan. O'zbek tili korpusi hozirgi kungacha to'liq yaratilmaganligi sababli bu masalani yechish dolzarb hisoblanadi. Korpusni tahlil qilish va yaratish uchun Python dasturlash tilida dasturiy ta'minot yaratilgan. Korpusni yaratish uchun turli xil janrlarga tegishli matnlar foydalanilgan.

Kalit so'zlar: Uzbek korpus, NLP, NLTK, XML. Uzbek tili.

Annotation: A language corpus is a collection of written or spoken texts that have been compiled and annotated for linguistic analysis. These corpora serve as a resource for researchers in fields such as linguistics, natural language processing, and computational linguistics. The creation of a corpus for the Uzbek language is a challenging task, given the limited digital resources and linguistic tools available for this language. In this paper, the theoretical and practical foundations of the creation of software for the national corpus of the Uzbek language are studied. Since the corpus of the Uzbek language has not yet been fully created, it is urgent to solve this issue. Software was created in the Python programming language to analyze and create the corpus. Texts of different genres were used to create the corpus.

Keywords: Uzbek corpus, NLP, NLTK, XML, Uzbek language.

Аннотация: Языковой корпус представляет собой набор письменных или устных текстов, которые были скомпилированы и аннотированы для лингвистического анализа. Эти корпуса служат ресурсом для исследователей в таких областях, как лингвистика, обработка естественного языка и компьютерная лингвистика. Создание корпуса для узбекского языка является сложной задачей, учитывая ограниченность цифровых ресурсов и лингвистических инструментов, доступных для этого языка. В данной статье изучены

теоретические и практические основы создания программного обеспечения национального корпуса узбекского языка. Поскольку корпус узбекского языка еще полностью не создан, необходимо срочно решить этот вопрос. Программное обеспечение было создано на языке программирования Python для анализа и создания корпуса. Для создания корпуса использовались тексты разных жанров.

Ключевые слова: Узбекский корпус, НЛП, НЛТК, XML, узбекский язык.

Tabiiy tillarni anglash (NLP: Natural Language Processing) sohasi boshqa sohalarga qaraganda o'zining dolzarbligi jihatidan ancha oldinlab ketdi va bu sohada barcha rivojlangan tillar qaotiri, o'zbek tili uchun ham bir qancha til resurslari va dasturiy vositalari ishlab chiqila boshlandi [E. Kuriyozov, U. Salaev, S. Matlatipov, and G. Matlatipov. 2023:3]. Korpus - tilni yoki tilning o'zgarishini to'liq ifodalash uchun tashqi mezonlarga muvofiq tanlangan elektron shaklda matn qismlari to'plamidir. U lingvistik tadqiqotlar uchun ma'lumot bazasi sifatida ishlatiladi.

Korpus – barcha uchun ochiq tarzda havola qilingan elektron matnlar majmuidir. Muayyan tilga doir elektron matnlar tizimga qancha ko'p joylashtirilsa, shuncha mazkur tilda turli fanlarga doir axborotlarni olish, kompyuter tarjimai uchun zarur bo'lgan lingvistik platformani yaratish imkoniyati paydo bo'ladi.

O'zbek olimlaridan B.Mengliyev, Sh.Shahobiddinova, Z.Xolmanova, S.Karimov, N.Abduraxmonova, L.Raupova, Sh.Hamroyeva, M.Abjalova, G.Toirova, G.Ikromova, J.Djumabayeva, G.Ergasheva, A.Eshmo'minovlar ilmiy ishlar qilishgan. O'zbek tili milliy korpusining konseptologiyasi B.Mengliyev rahbarligidagi olimlar jamoasi tomonidan ishlab chiqilmoqda.

Ushbu ilmiy ish natijasi o'zbek tili uchun yaratilayotgan NLP modellar va dasturiy vositalar qatorida yangi ilmiy izlanishlar uchun imkon yaratadi. Shu paytgacha NLP sohasida o'zbek tili uchun bir qancha ilmiy ishlar qilingan, shular qatoriga o'zbek tili uchun yaratilgan nomuhim so'zlar dataseti [4]–[6], lemmatizatsiya modellari [7], morfologik va sintaktik belgilangan dataset [8], hamda sentiment tahlili datasetlari [9] kiradi. Shu bilan birga, bir qancha NLP dasturiy vositalar ham o'zbek tili uchun yaratilgan bo'lib, Lotin-Kirill alifbosi transliteratsiyasi [10], stemmer [11] dasturlari shular jumlasidandir.

Biz korpus yaratish maqsadida turli xil janrlardagi matnlarni yig'ish uchun XML fayllaridan foydalandik. XML tili tilshunoslikka doir ma'lumotlarni saqlash va uzatish uchun qulay vosita hisoblanadi. Jumladan, XML teglarining foydalanuvchi tomonidan yaratilishi, yangi ma'lumotlar qo'shilganda alohida element sifatida qo'shib ketilishi, qidiruv jarayonlarining tez ishlashi bilan boshqa ma'lumotlar tuzilmalaridan ajralib turadi. Shuning uchun o'zbek tili korpusini va korpus haqidagi ma'lumotlarni saqlash uchun XML tilidan foydalanildi. Mazkur korpusni yaratish uchun Python dasturlash tili ishlatilgan, Python dasturlash tilining ElementTree moduli yordamida XML formatda saqlangan ma'lumotlarni qayta ishlash mumkin. Python dasturlash tilining NLTK (Natural Language Toolkit) kutubxonasida ham XML tilini ishlatish imkoniyati mavjud, bu esa tabiiy tillarni qayta ishlash masalalarini yechish uchun qulay hisoblanadi, shularni hisobga olgan holda XML tilidan foydalanib korpusda saqlanadigan matnlarni tasvirlash uchun quyidagi ko'rinishdagi metama'lumot teglari yaratildi:

<CORPUS NAME=" " > </CORPUS> Korpus haqidagi ma'lumotlar

<TEXT ID=" " TYPE=" " STYLE=" " > </TEXT> - Hujjat haqida ma'lumotlar

<AUTHOR> </AUTHOR> - Hujjatning mualliflari

<SOURCE> </SOURCE> - Hujjat olingan manba

<DATE> </DATE> - Hujjat yozilgan sana

<CATEGORY> </CATEGORY> - Hujjatning qaysi janrga tegishli ekanligi

<CONTENT TITLE=" " </CONTENT> - Hujjat matnini saqllovchi teg

Bu yerda NAME=' ' atributida korpus nomi, TYPE=' ' atributida matnning turi(hikoya, she'r, qissa va h.), STYLE=' ' atributida matnning uslubi hamda TITLE=' ' atributida har bir hujjatning sarlavhasi saqlanadi.

Lingvistik tuzilmalar uchun XML dan foydalanish: Moslashuvchanligi va kengayishi tufayli XML lingvistik tuzilmalarni ifodalash uchun tabiiy tanlovdir. Bu erda oddiy leksik yozuvga misol keltirilgan.

<?xml version='1.0' encoding='UTF-8'?>

<corpus name=" ">

<text id='1' type=" " style='publististik'>

<author></author>

<source>http://darakchi.uz/oz/115586</source>

<date>15.03.2021, 12:12</date>

<category>Sport</category>

<content title='Mamlakatimizda ilk bor paraqilichbozlik bo'yicha chempionat bo'lib o'tdi'> Barchamizga ma'lumki so'ngi yillarda barcha sohada bo'lgani kabi sport sohasida ham bir qator o'zgarishlar va ezgu ishlar amalga oshirilmogda. </content> </text> </corpus>

Dasturlashda qulaylik uchun XML ma'lumotlariga satr sifatida murojaat qildik.

Shu algoritm asosida yaratilayotgan o'zbek tili korpusining strukturasi va undagi so'zlar hamda gaplar soni quyidagi jadvalda keltirilgan:

Таблица 1.

Yaratilayotgan korpus tarkibidagi matlar sohalar kesimida statistikasi

Janrlar	Maqolalar	Gaplar	Jami so'zlar	Qatnashgan so'zlar (Unique words)
Texnika va texnologiya	819	9497	124655	25331
Iqtisodiyot	899	10009	164558	21985
Huquq	900	10890	176429	22480
Madaniyat	900	13806	207037	35741
Maqolalar va hikoyalar	586	56205	496045	83667
Salomatlik	900	15105	184687	31871
Shou-biznes	918	23539	232268	43270
Siyosat	899	12477	222832	26498
Sport	899	9810	136840	22957
Boshqa janrlar	34	130966	1942562	159348
Jami	7754	292304	3887913	261879

Jadvaldagi boshqa janrlar bo'limiga quyidagilar kiradi: diniy, badiiy asarlar, fan va ta'lim, qonun hujjatlari, tibbiyot, qishloq xo'jaligi.

Ko'rib turganimizdek, fayl korpus nomi, matn nomeri, stili, muallif, matn yaratilgan sana, va boshqalar uchun XML teglarini o'z ichiga oladi. Yaratilgan korpus Texnika va texnologiya,

Iqtisodiyot, Huquq, Madaniyat, Maqolalar va hikoyalar, Salomatlik, Shou-biznes, Siyosat, Sport, Boshqa janrlardan tashkil topgan. Shu jumladan jami maqolalar 7754 ta, gaplar soni 292304 ta, soʻzlar soni 3887913 ta, qatnashgan soʻzlar (Unique words) 261879 tani tashkil qiladi. Biz bu korpusni boyitishda davom qilyapmiz.

Foydalanilgan adabiyotlar:

1. E. Kuriyozov, U. Salaev, S. Matlatipov, and G. Matlatipov, "Text classification dataset and analysis for Uzbek language," *arXiv preprint arXiv:2302.14494*, 2023.
2. S. Matlatipov, H. Rahimboeva, J. Rajabov, and E. Kuriyozov, "Uzbek Sentiment Analysis Based on Local Restaurant Reviews," in *CEUR Workshop Proceedings*, 2022, pp. 126–136. [Online]. Available: www.scopus.com
3. M. Sharipov, E. Kuriyozov, O. Yuldashev, and O. Sobirov, "UzbekTagger: The rule-based POS tagger for Uzbek language," *arXiv preprint arXiv:2301.12711*, 2023.
4. K. Madatov, S. Bekchanov, and J. Vičič, "Accuracy of the Uzbek stop words detection: a case study on» School corpus»,," *arXiv preprint arXiv:2209.07053*, 2022.
5. K. Madatov, S. Bekchanov, and J. Vičič, "Dataset of stopwords extracted from Uzbek texts," *Data Brief*, vol. 43, p. 108351, 2022.
6. X. Madatov, M. Sharipov, and S. Bekchanov, "O'ZBEK TILI MATNLARIDAGI NOMUHIM SO'ZLAR," *COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS*, vol. 1, no. 1, 2021.
7. M. Sharipov and O. Sobirov, "Development of a rule-based lemmatization algorithm through Finite State Machine for Uzbek language," *arXiv preprint arXiv:2210.16006*, 2022.
8. M. Sharipov, J. Mattiev, J. Sobirov, and R. Baltayev, "Creating a morphological and syntactic tagged corpus for the Uzbek language," *arXiv preprint arXiv:2210.15234*, 2022.
9. E. Kuriyozov, S. Matlatipov, M. A. Alonso, and C. Gómez-Rodríguez, "Construction and evaluation of sentiment datasets for low-resource languages: The case of Uzbek," in *Human Language Technology. Challenges for Computer Science and Linguistics: 9th Language and Technology Conference, LTC 2019, Poznan, Poland, May 17–19, 2019, Revised Selected Papers*, 2022, pp. 232–243.
10. U. Salaev, E. Kuriyozov, and C. Gómez-Rodríguez, "A machine transliteration tool between Uzbek alphabets," in *CEUR Workshop Proceedings*, 2022, pp. 42–50. [Online]. Available: www.scopus.com
11. M. Sharipov and U. Salaev, "Uzbek affix finite state machine for stemming," *arXiv preprint arXiv:2205.10078*, 2022.

PARALLEL KORPUSLARNI YARATISH ASOSLARI

Abjalova Manzura AbdurashetovnaFilologiya fanlari doktori (DSc),
Alisher Navoiy nomidagi Toshkent davlat
o'zbek tili va adabiyoti universiteti dotsenti
E-mail: abjalova.manzura@gmail.com

Annotatsiya: Parallel korpuslar til korpuslarning maqsadiga va tillariga ko'ra alohida turi bo'lib, uning bazasida ikki til yoxud bir tilning ikki xil xususiyatiga xos matnlar jamlangan bo'ladi. Shu bois uning nomi ko'plikda qo'llaniladi, ya'ni parallel korpus emas, parallel korpuslar. Mazkur maqolada parallel korpuslarning kelib chiqishi va ulardan foydalanilish omillari, parallel korpuslarning ahamiyati borasida so'z yuritildi.

Kalit so'zlar: parallel korpuslar, bimatn, ikki til, asliy matn, tarjima matni.

Annotation: Parallel corpora are a special type of language corpora, depending on their purpose and languages, which are based on two languages or texts characteristic of two different characteristics of one language. Therefore, its name is used in the plural, that is, parallel corpora, not parallel corpus. In this article, the factors of the origin and use of parallel corpora, the meaning of parallel corpora were considered.

Keywords: parallel corpora, bitext, two languages, original text, translation text.

Аннотация: Параллельные корпуса – это особый тип языковых корпусов, в зависимости от их назначения и языков, которые основаны на двух языках или текстах, характерных для двух различных характеристик одного языка. Поэтому его название используется во множественном числе, то есть параллельные корпуса, а не параллельный корпус. В этой статье были рассмотрены факторы происхождения и использования параллельных корпусов, значение параллельных корпусов.

Ключевые слова: параллельные корпуса, битекст, два языка, оригинальный текст, текст перевода.

Ikki tilli Kanadada tarjima masalasi hamisha muammoli bo'lib kelgan. Shu bois kanadalik tadqiqotchilar tarjima matnlaridagi tarjima qilinadigan fragmentlargina olingan matnlardan foydalanishni taklif qilishgan [Копотев, 2014; 130]. Mana shu tarzda bir tildagi matnning turli tildagi tarjimalari mavjud matnlar korpusi yuzaga kelgan.

Rozettsk toshi – PK uchun dastlabki misol. U er.avv. II asrda yaratilgan bo'lib, ushbu tosh o'rta qismidagi demotik bir matn qadimgi Misr iyerogliflari (toshning tepa qismi) va qadimgi yunon yozuvi (toshning pastki qismida)da o'z aksini topgan.



Rozettsk toshi (er.avv. II asr)

Parallel korpuslar yohud **tarjimalar korpusi** (ingl. *parallel / translation corpus*) – bu asl tildagi matnlar va uning boshqa tildagi tarjimasi mavjud korpus. Qoidaga ko‘ra, bunday korpuslar faqat ikkita tilda tayyorlanadi, shu bois undagi o‘zaro mutanosib matnlar *bitekst* ham deyiladi. **Bimatn (bitekst)** – asliy til matni va uning boshqa til(lar)dagi tarjimasi mavjud matnlar.

Parallel korpuslar – partekst – bimatn – bitekst.

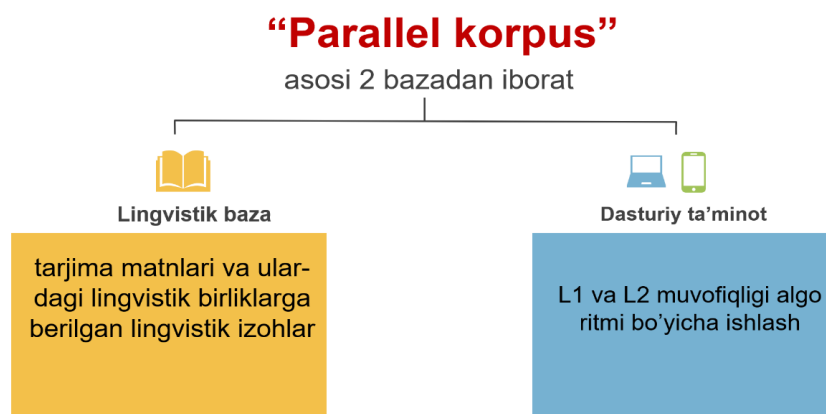
Bitekstni yaratish g‘oyasi Brayan Xarris (Brian Harris)ga tegishli, u 1988-yilda bitekstni yaratish konsepsiyasi bo‘yicha tadqiqot yozgan va keyinchalik Monreal universiteti (Université de Montréal)ning bir guruh olimlari tomonidan rivojlantirilgan va konsepsiya RALI (Recherche appliquée en linguistique informatique или Applied Research in Computational Linguistics – «Прикладные исследования в вычислительной лингвистике» – “Hisoblash tilshunosligida amaliy tadqiqotlar) deb nom olgan. Guruh tabiiy matnni qayta ishlashni o‘rgangan dasturchilar va tilshunoslardan iborat bo‘lgan. Bitext konsepsiyasining taniqli targ‘ibotchilari Per Isabelle (Pierre Isabelle) va Klod Bédard (Claude Bédard)dir.

Parallel korpuslar, asosan, ma‘lum bir kontekstdagi turli til tuzilmalari, iboralar va so‘zlarning tarjimasini ko‘rish uchun ishlatiladi.

Parallel korpuslar matnlarining tarjima yo‘nalishi bo‘yicha turlari mavjud:

- ✓ **bir yo‘nalishli** (masalan, inglizcha matn o‘zbek tiliga tarjima qilingan)
- ✓ **ikki tomonlama** (masalan, inglizcha matn o‘zbek tiliga tarjima qilingan va aksincha, ya‘ni ingliz ↔ o‘zbek)
- ✓ **ko‘p yo‘nalishli** (inglizcha matn o‘zbek, nemis, fransuz va boshqalarga tarjima qilingan)

Shu bois PK bir tilli (agar bir tilning bir nechta variantlari, shevalari qiyoslansa), ikki tilli va ko‘p tilli bo‘ladi.



Parallellik mezoniga ko‘ra, korpuslar ikkita asosiy turga bo‘linadi:

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.

Ikkala turdagi korpuslar ikki tilli va ko‘p tilli terminologik lug‘atlarni tuzishda, shuningdek, tillarni qiyosiy o‘rganishda (leksikologiya, grammatika, stilistika, dialektologiya, tarjimashunoslik va boshq.) samarali tarjima usullarini, shu jumladan, mashina tarjimasi tizimlarini ishlab chiqish uchun ishlatiladi.

Birinchi turdagi parallel korpuslarni tayyorlash va ularni qayta ishlash dasturlarini ishlab chiqishda *moslashtirish* (alignment) – asl matn bilan tarjima matnlari fragmentlari o‘rtasidagi muvofiqlikni o‘rnatish muammosi yuzaga keladi. Ushbu muammoni hal qilish uchun matnlarni avtomatik ravishda to‘g‘rilashning turli usullari qo‘llaniladi: jumlar bo‘yicha (jadval. 4.2), klauzlar (grammatik tuzilmalar) bo‘yicha, so‘z birikmalari va so‘zlar [Захаров, Богданова, 2020; 61].

Ta’kidlash joizki, bugungi kunda mashina tarjimasida so‘zma-so‘z tarjima usulidan voz kechilgan bo‘lib, ushbu usul hatto an’anaviy tarjima jarayonida ham o‘zini oqlamagan. Hatto asl matndagi katta bir xatboshi bir necha gap bilan tarjima matnda o‘z mazmunini topgan. Shu bois parallel korpuslarda ham tarjimalar jarayonida jumlama-jumla moslashish bo‘lmasligi mumkin. Shu jihadni e’tiborga olib, Zubovalar darsligida asl va tarjima matnlarida gaplarni to‘g‘rilashning 6 xil ehtimoliy muvofiqliklari uchrashi ta’kidlanadi. Ular quyidagilar:

- 1) asl matndagi bitta gap aynan bitta gap bilan tarjima qilinadi;
- 2) ikkita gap bitta gap bilan tarjima qilinadi;
- 3) asl matndagi ikkita gap bitta gapga tenglashtirilib tarjima qilinadi;
- 4) ikkita gap bitta gap bilan tarjima qilinadi, ammo gaplarning ichki chegarasi asliy matndagi bilan tarjima matndagi gaplarga mos kelmaydi;
- 5) asl matndagi gap tarjima qilinmaydi, chunki mazmunan buning hojati bo‘lmaydi;
- 6) tarjima matnidagi gap asliy matnda bo‘lmasligi mumkin [Зубов, Зубова 2004].

Tarjima jarayoni nozik ish hisoblanadi. Asliy matndagi mazmun tarjima matnida o‘z ifodasini topishi kerak. Muayyan qo‘shimcha boshqa tilda bir necha qo‘shimcha yoki predloglarga mos kelishi mumkin. Masalan, -ning qaratqich kelishigi ingliz tilida *’s* va *of* predlogi bilan, *-ga*, *-gacha*, *uchun* kabi grammatik ifodalar *to* predlogi yordamida tarjima qilinishi mumkin. Bunday klauzlar bilan birga so‘zlarda ham shunday muammo uchraydi. Masalan, ingliz tilidagi *like* so‘zi *yoqtirmoq* va *kabi*, *-dek*, *o‘xshab* birliklarning tarjimasi bo‘lib keladi.

PKni yaratish uchun asliy til va uning tarjimasining mavjudligi yetarli emas. Asosiy masala asliy tildagi fragmentning tarjima tilda o‘z ifodasini topishi hisoblanadi. Mana shu maqsadda moslashtirish (ingl. *alignment*) jarayoni qo‘llaniladi. Natijada parallel matnlardagi bir xil fragmentlar bir-biriga qiyosan taqdim etiladi [Abjalova, 2022].

Muvofiqlashtirishning quyidagi turlari mavjud:

– *so‘zma-so‘z muvofiqlashtirish* (word alignment) – eng ideal, ammo eng katta muammo tillardagi leksemalar, frazemalar, so‘zshakllari har doim ham o‘z muqobiliga ega bo‘lavermaydi. Dastlabki mashina tarjimalari so‘zma-so‘z tarjimani amalga oshirgan.

- *gaplarni muvofiqlashtirish* (sentence alignment) – nisbatan samarali usul.
- *abzas bo'yicha muvofiqlashtirish* (paragraph alignment).

Muvofiqlashtirish masalasi bilan birga yana bir muammo ko'ndalang bo'ladi: bimatnlarning o'zaro mos bo'lmay qolishi. Tarjimonlar senzura yo'l qo'ymaslik, matnlarni qisqartirish, bir qancha gapni tarjimada bitta gap bilan ifodalash, tarjimaga ijodiy yondashish yoki psixologik omillar sabab mazmunni chiqarib bera olmaslik kabi sabablar tarjima matnidagi mazmunning original matn mazmuniga mos tushmasligi mumkin.

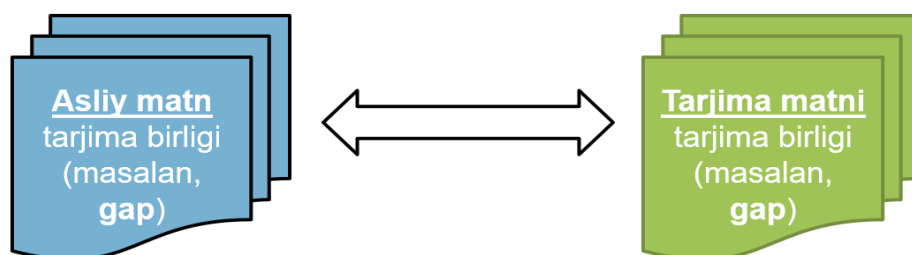
Shunisi aniqki, PK matnni qanday tarjima qilishni o'rgatmaydi, balki tarjimadagi o'zaro mosliklarni ko'rsatadi, ya'ni muayyan leksema, so'z shakl yoki grammatik formaga boshqa tilda unga muvofiqlashtirilgan birlik konkordansda taqdim etiladi.

Matnlarni muvofiqlashtirishning bir qancha dasturlari mavjud. Ular muayyan uzunlikdagi gaplarni avtomatik ravishda o'zaro qiyoslaydi, matnni paragraflarga bo'ladi, tinish belgilarini tahlil qiladi, lug'atlar bilan muvofiqlashtiradi va boshq. Ko'p hollarda ushbu dasturlar dialog rejimida yoki avtomatik moslashtirish natijalarini tahrirlashdan keyin inson-mashina holatida qo'llaniladi. Bunday dasturlarga misol tariqasida quyidagilarni aytish mumkin: Humaling, Abbyy Aligner, Trados, Winalign, Wordfast tools, Giza++ va boshq.

Kelajakda PKdagi eng asosiy masala parallel lingvistik izoh (teg, razmetka)larning muvofiqlashishi hisoblanadi. Bunda turli tillardagi grammatik va leksik muvofiqliklar avtomat tarzda topiladi.

“Слова о полку Игореве” asarining rus tiliga yuzdan ortiq tarjimasi, ikki yuzdan ortiq boshqa tillarga, jumladan, o'zbek tiliga ham tarjimasi mavjud [<http://nevmenandr.net/slovo/>]. Ushbu asar boshqa tillarga gap muvofiqligi bo'yicha tarjima qilingan. Shu bois gapdagi ayrim so'zlar boshqa so'zlar bilan o'z ifodasiga ega bo'lgan. Masalan, «растекашется мысляю по древу» yoyiq birikmasidagi *мысль* so'zi *соловей*, *мысль*, *векаша*, *белка* kabi tarjimaga ega bo'lgan [Копотев, 2014; 134].

Ko'p tilli korpuslarni moslashtirish Aligned multilingual corpora



Korpusdagi tegishli tarjima birliklari bog'langan, ya'ni matnni uning tarjimasi bilan moslashtirish – bu matnning muayyan qismi tarjima matnining qaysi qismiga mos kelishini ko'rsatishdir.

Masalan:

Search TextBases

faculté de

Options

☒ Consecutive Words

☐ Partial Match(es)

☐ Stemmed Search

☐ Fuzzy

☒ Search Latest Content First

Search Results:

Item Select	Expression	Frequency	Source
1	faculté de	144	0

Item Select	#	Sources	Sentence	Word
Select	15	A_CN4_0574_F 1053	9	
Select	16	A_CN4_0574_F 1310	9	
Select	17	A_CN4_0574_F 1954	10	
Select	18	A_CN4_0573_F 1274	6	
Select	19	A_CN4_0566_F 9	14	
Select	20	A_CN4_0566_F 669	36	
Select	21	A_CN4_0566_F 778	43	
Select	22	A_CN4_0565_F 1295	22	

Source [French] - A_CN4_0573_F.inp

118 - 1 Le même code de la nationalité camerounaise dispose en son article 31 : « Perd la nationalité camerounaise :

a) Le Camerounais majeur qui acquiert ou conserve volontairement une nationalité étrangère;

b) Celui qui exerce la **faculté de répudier la qualité de camerounais conformément aux dispositions de la présente loi;**

c) Celui qui, remplissant un emploi dans un service public d'un organisme international ou étranger, le conserve nonobstant l'injonction de le résigner faite par le Gouvernement camerounais ».

119 - 1 Permanent Court of Arbitration, Eritrea c. Ethiopia Claims Commission, Partial Award Civilians Claims, Eritrea's Claims 15, 16, 23 & 27-32, between the State of Eritrea and the Federal Democratic Republic of Ethiopia, The Hague, December 17, 2004, par. 10.

120 - - Voir Partial Award Civilians Claims, Eritrea's Claims 15, 16

Target [English] - A_CN4_0573_E.inp

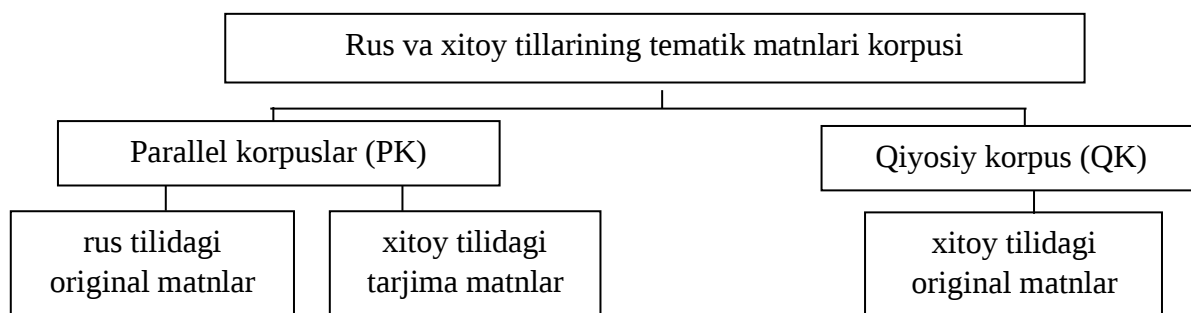
"The following shall lose Cameroonian nationality:

(a) A Cameroonian who, having reached the age of majority, voluntarily acquires or retains a foreign nationality;

(b) A person who exercises the **option to repudiate his status as a Cameroonian according to the provisions of this Act;**

(c) A person employed in the public service of an international or foreign entity who remains in that employment despite an order to resign from it issued by the Government of Cameroon."

Parallel korpuslar, asosan, quyidagi strukturaga bo'linadi (rus va xitoy tillarining tematik korpusi strukturasi asosida tushuntirilmoqda):



PK matnlari yordamida katta hajmdagi ma'lumotlarga ega bo'lish mumkin. Ular quyidagilar bilan belgilanadi:

- ikki tilli va ko'p tilli tarjima lug'atlari yaratish;
- mashina tarjimasi tizimlari lug'atlarini yaratish va to'ldirish mumkin;
- kontekstni kompyuter yordamida tahlil qilish orqali ko'p ma'noli so'zning kontekstual bir ma'nosini qoldirish, bunda kontekst uzunligi bo'yicha gapdan kengroq bo'lishi mumkin;
- matnning terminologik va frazeologik birliklarini tarjima qilish;
- semantik maydonlar va terminologik tizimlarni formallantirish;
- tarjima universallarini o'rganish;
- kompyuter xotirasida turli darajalarda bir-biriga mos keladigan asliy til korpuslari matnlari va ularning tarjimalarini to'plash orqali tarjima xotirasiga ega yangi mashina tarjima tizimlari doirasida to'liq avtomatik tarjimani amalga oshirish.

Tarjima jarayonida bunday tizim tarjima qilingan gap yoki uning parchasini asl parallel matnlar qatorida topishga harakat qiladi. Agar u asl matnlardan topilgan bo'lsa, unda tizim bunday gapning tarjimasi yoki uning parchasini tarjima qilingan matnlar qatoridan tanlaydi.

Parallel korpuslarga namunalari:

- Glosbe: <https://glosbe.com/>
- Linguee: <https://www.linguee.ru/>
- MyMemory: <https://mymemory.translated.net>
- Opus: <https://www.opus.lingfil.uu.se>
- Reverso: <https://www.context.reverso.net>
- 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.^[1]
- [European Parliament Proceedings Parallel Corpus 1996–2011](#)
- [The Opus project aims at collecting freely available parallel corpora](#)
- [Japanese-English Bilingual Corpus of Wikipedia's Kyoto Articles](#)
- [COMPARA – Portuguese/English parallel corpora](#)
- [TERMSEARCH – English/Russian/French parallel corpora \(Major international treaties, conventions, agreements, etc.\)](#)
- [TradooIT – English/French/Spanish – Free Online tools](#)
- [Nunavut Hansard – English/Inuktitut parallel corpus](#)
- [ParaSol – A parallel corpus of Slavic and other languages](#)
- [Glosbe: Multilanguage parallel corpora](#) with online search interface
- [InterCorp: A multilingual parallel corpus](#) 40 languages aligned with Czech, [online search interface](#)
- [myCAT – Olanto](#), concordancer (open source AGPL) with online search on JCR and UNO corpus
- [TAUS](#), with online search interface.
- [linguatools](#) multilingual parallel corpora, online search interface.
- [EUR-Lex Corpus – corpus](#) built up of the [EUR-Lex](#) database consists of [European Union law](#) and other public documents of the [European Union](#)
- [Language Grid – Multilingual service platform that includes parallel text services](#)

Muvofiqlashtirish dasturlari:

- [GIZA++ alignment tool \(1999\)](#)
- [Uplug – tools for processing parallel corpora \(2003\)](#)
- [An implementation of the Gale and Church sentence alignment algorithm \(2005\)](#)
- [The Hunalign sentence aligner \(2005\)](#)
- [Champollion \(2006\)](#)
- [mALIGNa \(2008–2020\)](#)
- [Gargantua sentence aligner \(2010\)](#)
- [Bleualign – machine translation based sentence alignment \(2010\)](#)
- [YASA \(2013\)](#)

- [Hierarchical alignment tool \(HAT\) \(2018\)](#)
- [Vecalign sentence alignment algorithm \(2019\)](#)
- [Web Alignment Tool at University of Grenoble](#)

Foydalanilgan adabiyotlar:

1. Abjalova M. Korpus lingvistikasi. [Matn]: uslubiy qo‘llanma / M.A. Abjalova. –Toshkent: Nodirabegim, 2022. – 110 b.
2. Захаров В., Богданова С. Корпусная лингвистика: учебник. 3-е изд., перераб. – СПб.: Изд-во С.-Петерб. ун-та, 2020. – 234 с.
3. Зубов А. В., Зубова И. И. Информационные технологии в лингвистике: учеб. пос. М.: Издательский центр «Академия», 2004.
4. Копотев М. Введение в корпусную лингвистику. – Прага, 2014. – 230 с.
5. <http://nevmenandr.net/slovo/>

DISCUSSION OF NANO-PREFIXED TERMS

Nurmatova Guzal Hakimovna

Doctor of Philosophy (PhD) in Philological Sciences,
Associate professor of Navoi Profi University

Abstract: This study introduces three problems of nano-prefixed terms. Firstly, the scientific and linguistic evolution of the prefix ‘nano’ is discussed. Secondly, views on collocation relations of the Eastern and Western scholars are compared and new terms for ‘node word’ and ‘collocated word’ are introduced as ‘orbital’ and ‘satellite’ terms in the terminology of Uzbek corpus linguistics. Thirdly, the characteristics of nano-prefixed terms have been analyzed in a corpus.

Keywords: nano-prefix, evolution, collocation, relation, node word, collocate word, orbital word, satellite word.

Анотация: В данном исследовании представлены три проблемы терминов с нано (нано) префиксом. Во-первых, обсуждается научная и лингвистическая эволюция приставки “нано”. Во-вторых, сопоставляются взгляды восточных и западных ученых на отношения словосочетаний и вводятся новые термины для “node word” и “collocate word” как “орбитальное слово” (orbital so‘z) и “слова-сателлиты” (yo‘ldosh so‘z) в качестве терминов в терминологии узбекской корпусной лингвистики. В-третьих, в корпусе проанализированы особенности нано префиксных терминов.

Ключевые слова: нано префикс, эволюция, словосочетание, зависимость, узловое слово, словосочетание, орбитальное слово, слово-сателлит.

One of the most promising technologies of the 21st century, developed on the basis of the effective application of the theory of nanoscience, is nanotechnology, which is currently being used in natural and artificial research in almost all fields as a rapidly developing fundamental and applied science.

Although the terms of nanotechnology is related with the name of Norio Taniguchi, a Japanese scientist, and nanoscience with the name of Feynman, an American scientist, some experts and linguists have expressed their opinions on the problems of distinguishing the terms nanotechnology and nanoscience, the origin, development and definition of these two concepts are different [Bayda, 2013:19].

First of all, if you look at the prefix “nano” of these terms, it means «dwarf» (deaf) or something very small from the Greek and describes one thousand millionth of a meter (10^{-9} m). Nanoscience is the study of structures and molecules at the nanometer scale, between 1 and 100 nm, and the technology that uses it in practical applications, such as devices [Mansoori, 2020:93]. In the International System of Units², the prefix “nano” is one billionth, or 10^{-9} m; therefore, one nanometer is one billionth of a meter [Gnach, 2020:112].

² (SI) (fr. Le Système International d'Unités (SI)) - system of units of physical quantities, a modern version of the metric system. SI defines seven basic and derived units of physical quantities (hereafter referred to as units), as well as a set of prefixes. SI defines seven basic and derived units of physical quantities (hereafter referred to as units), as well as a set of prefixes.

The prefix “nano” is a reduction of the original value by a billion times, that is, a unit of measurement consisting of nine zeros divided by one - 1,000,000,000. For example, 1 nanometer is one billionth of a meter. ($1 \text{ nm} = 10^{-9} \text{ m}$). L.B. Piotrovsky and E. A. Katsl proposed the following experiment to visualize how small 1 nm is (see Fig.1). They mean that if we reduce the diameter of our planet ($12,750 \text{ km} = 12.75 \times 10^6 \text{ m} \approx 10^7 \text{ m}$) by 100 million (10^8) times, we get about 10^{-1} m . This is roughly the size of a soccer ball (the standard diameter of a soccer ball is 22 cm, but this scale difference in our imagination is insignificant; for us $2,2 \times 10^{-1} \text{ m} \approx 10^{-1} \text{ m}$). If the diameter of a soccer ball is reduced to the same 100 million (10^8) times, then one can only imagine a nanoparticle size of 1 nm, roughly the diameter of a C_{60} fullerene carbon molecule, shaped like a soccer ball. (see Fig.1).

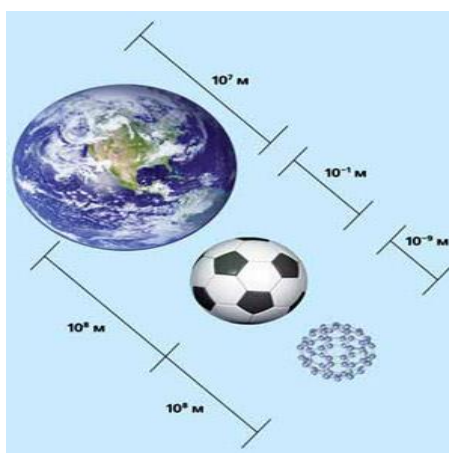


Fig 1. The ratio of the diameters of the Earth ($\approx 10^7 \text{ m}$), a soccer ball ($\approx 10^{-1} \text{ m}$) and a C_{60} molecule ($\approx 10^{-9} \text{ m} = 1 \text{ nm}$).

It is noteworthy that the prefix “nano” in the scientific literature does not refer to the field of nanotechnology, but to objects whose size is billions of times larger than 1 nm - in the terminology of dinosaurs, in particular, sizes of 5 and 1.3 m, respectively - nanotyrannosaurs (nanotyrannus) and nanosaurus (nanosaurus) - used for dwarf dinosaurs called dwarf dinosaurs. But their size was really “dwarf” compared to other dinosaurs, which exceeded 10 m (up to 50 m) and could have weighed 30-40 tons or more. This example highlights that the prefix «nano» itself has no physical meaning, only a measurement [Piotrovskiy, 2010].

Today this prefix marks a new fourth period of technological development - the era of nanotechnology, which is considered the industrial revolution.

The second issue of our research, the linguistic terms and phenomena that cause debates around “collocation” and “collocational links” by Western linguists and “words relations” in “The Origin of Sciences” by the Uzbek philosopher Abu Nasr Farabi. We put forward our hypothesis that the information about the “laws of words relations” today may be about collocational bonds and tried to substantiate this hypothesis. It should be noted that in our previous research, we mentioned the contribution of Farabi in terms of collocations [Nurmatova, 2021].

First of all, Farabi interpreted the problems of words relation as a separate science and divided it into several categories [Nurmonov, 2002:12].

By the category of words relations Abu Nasr Farabi, in our opinion, means not only dominant word and the subordinate words, but also other, before and after a word relation as well as the

combination of word parts. Such rules of words relations in corpus linguistics are called right and left collocational co-occurrence. In corpus linguistics the *node* word and the words that form a combination around it are the creation of associations. According to the rules of word combinations in the research conducted by the British linguist V. Brezina, the node word is intended to form a word or words that can be combined around it to form a completely different type of word combination in his research. V. Brezina's experiment helps to study the relation of words, as it was studied by Abu Nasr Farabi (see fig. 2) [Brezina et al, 2015:151].

These comparative aspects were discussed by G.Kh. o Nurmatova's previous research, here the node word and collocate words are compared to the English language terms *orbital word* and its *satellite words* and for the first time these terms were applied for Uzbek corpus linguistics' terminology [Nurmatova, 2021].

As an example, figure 2 and table 1 show the visualization and statistics of the collocation of the orbital prefix “nano” with satellite words on a specific device:

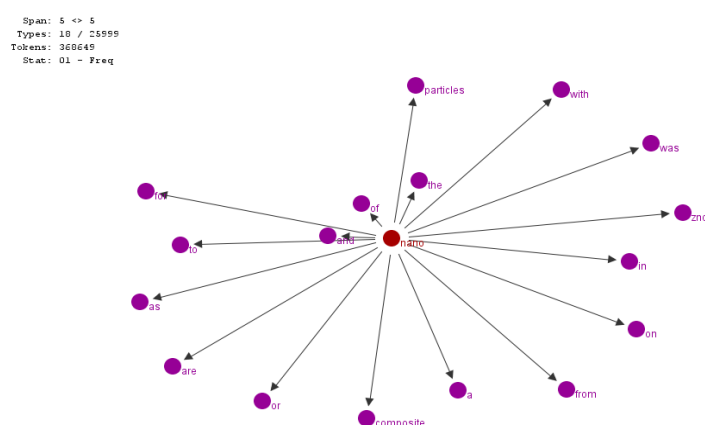


Figure 2. Collocational visualization of the *nano* prefix

Table 1.

Statistical table of the collocation of the *nano* prefix

Word type	Statistic	Freq.(within)	Word type	Statistic	Freq.(within)
and	45.0	45	as	9.0	9
of	38.0	38	on	7.0	7
the	32.0	32	or	6.0	6
to	17.0	17	with	6.0	6
in	13.0	13	zno	6.0	6
particles	12.0	12	composite	6.0	6
a	11.0	11	was	5.0	5
for	11.0	11	are	5.0	5

As it is seen, right and left collocation words in visual and tabular forms, here, the prefix *nano* indicates the connection of seventeen lexical units, indicating that it enters with different lexical units from the right and left sides. The *nano* prefix forms collocations not only with the

words *particles* and *composite*, i.e. *nanoparticles* and *nanocomposite*, but also with functional lexical units and their frequency.

P. Bayker, A. Hardy, McEnery divide the collocation of words in the corpus into two types: 1. *upward collocations*; and 2. *downward collocations*

When the frequency of the word orbital is lower than the frequency of the companion word, *upward collocations* are associated mainly with functional words, and such word associations are grammatical words, which can also be called *colligations*. [Baker, 2006:63, 165]. According to K. Brown and G. Miller, colligations, unlike collocations, refer to common lexical units connected by grammatical constructions (connections such as *in nano*, *are nano*, or *nano*). [Brown, 2015:86].

If the frequency of the orbital word is higher than the frequency of the companion word and they are mainly associated with context words, then the orbital word is contextually significant and can be considered a field term. For example, *nano* → *nanoparticles* → *nanocomposite* shows *downward collocations* in the semantic connection of collocations. It follows that there is a reason to call *nanoparticles* and *nanocomposite* collocations as terminological collocations.

References:

1. Bayda S., Adeel M., Tuccinardi T., Cordani M., Rizzolio F.. Nanotechnology: Interdisciplinary science of applications. African Journal of Biotechnology Vol. 12(3), pp. 219-226, 16 January, 2013. Available online at <http://www.academicjournals.org/AJB>
2. Brezina, V., McEnery, T. & Wattam, S. 2015. Collocations in context: A new perspective on collocation networks // International Journal of Corpus Linguistics, 20(2). -P. 151-157.
3. Baker P., Hardie A., McEnery T. A Glossary of Corpus Linguistics. Edinburgh: Edinburgh University Press, 2006. P. 192. –Pp. 63, 165.
4. Brown K. & Miller J.. Cambridge Dictionary of Linguistics. Cambridge University Press. 2015. -P. 86
5. Cotter M.J. Teaching terms: a corpus-based approach to terminology in ESP classes. Lisbon School of Accountancy and Administration (ISCAL) (Portugal). Pp. 499-506. P.501.
6. Gnach A., Lipinski T., in The History of Nanoscience and Nanotechnology: From Chemical–Physical Applications to Nanomedicine. 2020 Jan; 25(1): 112.
7. Марчук Ю.Н. Компьютерная лингвистика. –М.: Восток-Запад, 2007. –317 с. 93-99.
8. Nurmatova G.X. Problems on Corpus-Based Linguo-Statistic Study of Engineering Terms with application of AntConc and GraphColl Software Tools. Dissertation for degree of Doctor of Philosophy (PhD) in Philological Sciences. –Fergana State University, 2021. Pp. 193.
9. Nurmonov A. O‘zbek tilshunosligi tarixi: Oliy o‘quv yurtlari o‘zbek filologiyasi fakultetlarining talabalari uchun o‘quv qo‘llanma O‘quv qo‘llanma. –T.: O‘zbekiston. 2002. 232 b. B.11-12.
10. Пиотровский Л.Б., Кац Л.Б. «Нанотехнология», «Нанонаука» и «нанообъекты»: что значит «нано»?.. «Экология и жизнь» №8, №9 2010. https://elementy.ru/nauchno-populyarnaya_biblioteka

KORPUS LINGVISTIKASINING PROGRESSIV MASALALARI

Toshpo'latova Nilufar

O'zbekiston Milliy universiteti o'qituvchisi

Annotatsiya: Keyingi yillarda o'zbek tilshunosligi innovatsion yo'nalishda o'rganila boshlandi. Natijada o'zbek tilshunosligida kognitiv tilshunoslik, funksional grammatika, kommunikativ grammatika, korpus lingvistikasi va boshqa zamonaviy yo'nalishlar rivojlanmoqda. Faol madaniy-iqtisodiy muloqot va integratsiyaga asoslangan globallashuv jarayoni sharoitida axborot texnologiyalarining jadal rivojlanishi jamiyatdagi innovatsiyalar bilan bir qatorda har qanday fan sohasini ham yangilash va rivojlantirishni taqozo etmoqda. Bunday innovatsion rivojlanish jarayonini tilshunoslikda ham kuzatish mumkin. So'nggi o'n yil ichida kompyuter texnologiyalari imkoniyatlaridan foydalanish natijasida o'zbek tili ta'limida kompyuter tilshunosligi sohasi kengayib bormoqda. Tilshunoslikning kompyuter bilan uyg'unlashuvi tilshunoslikka yanada yangi imkoniyatlarni berdi hamda ayrim masalalarni kompyuter usullari yordamida avtomatlashtirilgan holda hal qilish imkoniyatlar yaratadi.

Kalit so'zlar: korpus, lingvistika, milliy korpus, mashina tarjimasi, tabiiy til, ma'lumotlar bazasi.

Аннотация: В последующие годы узбекское языкознание стало изучаться в инновационном направлении. В результате в узбекской лингвистике развиваются когнитивная лингвистика, функциональная грамматика, коммуникативная грамматика, корпусная лингвистика и другие современные направления. Бурное развитие информационных технологий в условиях процесса глобализации, основанного на активном культурно-экономическом диалоге и интеграции, наряду с инновациями в обществе требует обновления и развития любой отрасли науки. Такой процесс инновационного развития можно наблюдать и в лингвистике. В последнее десятилетие в результате использования возможностей компьютерных технологий в образовании узбекского языка расширяется область компьютерной лингвистики. Интеграция языкознания с компьютером дала лингвистике новые возможности, а также дает возможность решать некоторые вопросы автоматизированным способом с помощью компьютерных методов.

Ключевые слова: корпус, Лингвистика, Национальный корпус, машинный перевод, естественный язык, база данных.

Annonation: In the following years, Uzbek linguistics began to be studied in an innovative direction. As a result, cognitive linguistics, functional grammar, communicative grammar, corpus linguistics and other modern directions are developing in Uzbek linguistics. In the context of an active cultural and economic dialogue and a globalization process based on integration, the rapid development of Information Technology assumes the renewal and development of any field of science, as well as innovations in society. The process of such innovative development can also be observed in linguistics. In the past decade, the field of computer linguistics in Uzbek language education has been expanding as a result of the use of computer technology capabilities. The

combination of linguistics with computers has given linguistics more new opportunities and provides opportunities for solving certain issues in an automated way using computer methods.

Keywords: corpus, linguistics, National Corpus, machine translation, natural language, database.

Bugungi kunda texnikaning jadal rivojlanishi tilshunoslikda yangi yo‘nalishlarning paydo bo‘lishiga olib keldi. Ana shunday sohalaridan biri amaliy tilshunoslik doirasida vujudga kelgan korpus lingvistika fanidir. “Korpus”- kompyuter xotirasiga elektron tarzda asoslangan leksik va grammatik birliklarni to‘liq qamrab olgan matnlar to‘plami. Korpusning ahamiyati uni matnning annotatsiyasidadir, buning asosida so‘zlardan foydalanishning kommunikativ-pragmatik funksiyasini belgilaydigan matnlarni tahlil qilish maqsadi qo‘yiladi. Korpusga kiritilgan matnlar turli janrlarni (badiiy adabiy til, publitsistika, ilmiy-texnik matnlar va boshqalar) o‘z ichiga oladi. Matnlarni tanlash kompyuterning semantik tahlil qilish qobiliyatidan kelib chiqqan holda amalga oshiriladi. Ya’ni, ushbu tizimda tadqiqot ko‘laminin miqdoriy tabiati va sifat qiymatining uzluksizligi. Kompyuter texnikasining ko‘p qirraliligi matn lingvistikasi va matn tahlilining jadal rivojlanishi uchun katta imkoniyat yaratadi. Bu fanning paydo bo‘lishi hamda axborot tizimi va internetning globallashuvi til ta’limiga katta o‘zgarishlar olib keldi. “Korpus lingvistika” fan sohasi sifatida o‘zbek tilshunosligiga yaqinda kirib keldi. O‘zbek tili ta’limida korpus lingvistikasining dastlabki nazariy asoslari, milliy adabiy matnlar korpusining kelib chiqish tushunchalari, maqsad va vazifalari hamda ularga qo‘yiladigan talablar va boshqa masalalari bugungi kunda ilmiy muhokama qilinmoqda.

Til tadqiqotlarining zamonaviy yo‘nalishlari va axborot texnologiyalarining murakkab tabiati asosida aniqlangan tilning yangi kommunikativ darajasi bilan bevosita bog‘liq bo‘lgan ushbu yo‘nalishdagi asarlarning rivojlanishi tilshunoslikda yangi yo‘nalishlarning shakllanishiga ta’sir ko‘rsatdi. Ana shunday yangi yo‘nalishlardan biri korpus lingvistikasidir. Bugungi kunda texnika va kompyuterlarning jadal rivojlanishi tufayli shakllanayotgan korpus lingvistikasining nazariy ahamiyati va amaliy ahamiyati, eng avvalo, tadqiqotchilarga barcha ma’lumotlarni kompyuterga oldindan kiritish orqali kerakli ma’lumotlarni to‘plash imkonini beradi. Kompyuter texnikasi imkoniyatlaridan kelib chiqib, xalq ma’naviy va moddiy madaniyatining yorqin ko‘zgusi bo‘lmish tilshunoslikni rivojlantirishning ushbu yo‘nalishi til xazinasini har tomonlama keng tadqiq etish, turli saviyalarda tahlil qilish, innovatsion usullardan foydalanishdir. Shu munosabat bilan adabiy so‘zlar tilini korpus tilshunosligi sohasida keng yo‘nalishda o‘rganish juda muhimdir. Buning sababi shundaki, u o‘z navbatida o‘rganilayotgan matnlarda kompyuter imkoniyatlari orqali adabiy faoliyatning keng kanalini birlashtirdi. Shu munosabat bilan zamonaviy tilshunoslik tadqiqotlarida tildan foydalanish, yuqorida aytib o‘tganimizdek, tilning kommunikativ faoliyatining yangi darajasi va yangi axborot ijtimoiy ongining shakllanishi natijasida ulardan foydalanish xususiyatlariga qaratilgan. Bu holat tilni nafaqat o‘z-o‘zidan, balki til-kompyuter-inson majmuasida ham har tomonlama ko‘rib chiqishni taqozo etadi. Shu munosabat bilan, korpus matnini yaratishga qaratilgan maxsus ish, birinchi navbatda, matnlarni tanlashdan boshlanadi. Korpusga kiritilgan matnlarning tadqiqot maqsadi va obyektining ehtiyojlarini to‘liq qondirishi juda muhim bo‘lganligi sababli, hajmli matnlar tanlanadi. “Korpus” kompyuter xotirasiga elektron tarzda kiritilgan matnlar yig‘indisi desak, korpus lingvistikasining keng yo‘nalishidagi tadqiqotlar kompyuter imkoniyatlaridan kelib chiqqan holda katta hajmdagi elektron matnlar yordamida amalga oshiriladi. Shuning uchun korpus lingvistikasi bilan matn lingvistikasining yaqin aloqasiga

e'tibor berish juda muhimdir. Ma'lumki, so'nggi paytlarda matn tilshunosligi bilan bog'liq izlanishlar jadal olib borilmoqda. "Korpus lingvistika" fanining matn tanlashda xalqimizning milliy qadriyatlari, madaniyati va turmush tarzidan ma'lumot bera oladigan, tilshunoslik ma'lumotlarini ochib beruvchi asarlarni tanlab olishi kognitiv va informatsion ahamiyatga ega. Binobarin, korpusdan nafaqat katta hajmdagi matnlarni jamlovchi fond, balki merosimizni kelajak avlodlarga yetkazish vositasi sifatida ham samarali foydalanish zamonaviy o'zbek tili ta'limi yo'lidagi g'ayratli faoliyatdir.

Keyingi yillarda "Korpus lingvistika" fan sohasi sifatida yaqqol ustuvor o'ringa ega bo'ldi. Chunki bu sohada olib borilgan izlanishlar natijalari – matn korpusidan foydalanmasdan, ayniqsa, kompilyatsiya qilishdan foydalanmasdan til tadqiqotida tajriba o'tkazish oson emasligini ko'rsatdi.

Ma'lumki, tilshunoslik tadqiqotlarining asosiy qismi kompyuterlashtirilgan matn korpusini lingvistik tahlil qilishga qaratilgan. Korpus lingvistikasining yutuqlaridan biri sifatida korpusning eng rivojlangan turi – Milliy korpusni qayd etish mumkin. U qaysidir ma'noda Milliy tilni butunligicha ifodalaydi. Demak, Milliy korpus har bir tilning lingvistik qonuniyatlarini, jumladan, muntazam leksik-grammatik tizimini to'liq qamrab olgan (chuqur izohlangan) millionlab so'z qo'llanishlaridan iborat elektron shakldagi hajmli matnlar to'plamidir.

Milliy korpus, birinchi navbatda, tadqiqotchilarga lug'at va grammatikani har tomonlama o'rganish imkonini beradi. Albatta, tillarni ona tili yoki chet tili sifatida o'qitishda Milliy korpusdan foydalanishning ahamiyati ham ortib bormoqda. Shuning uchun ham hozirgi kunda ko'plab darsliklar va o'quv dasturlari matn korpusi asosida tuzilmoqda. Masalan, xorij fuqarolari, talabalar, o'qituvchilar, jurnalistlar, yozuvchilar elektron korpus yordamida noma'lum ma'noli so'z va grammatik shakllarning qo'llanish xususiyatlarini tez va samarali o'rganishi yoki tekshirishi mumkin.

Korpus lingvistikasi tilshunoslikning alohida sohasi sifatida matematik tilshunoslik, diskurs tahlili va leksikografiya sohalari bilan birlashtirilgan. Korpus lingvistikasining til fanining boshqa predmetlari bilan yaqinligi, bir tomondan, matnlar korpusining korpus lingvistika faoliyati natijasiga, ikkinchi tomondan, boshqa turlar uchun asosiy empirik material bo'lish imkoniyatiga bog'liq.

Milliy korpus leksikografiya, sun'iy intellekt, adabiyotshunoslik, og'zaki nutqni tahlil qilish va yig'ish hamda tilshunoslikning barcha sohalari bilan bog'liq tadqiqotlarni ta'minlaydi. Shu bilan birga, mashhur akademik lug'atlarni tuzish va ilmiy grammatikalarni yozishda korpuslar alohida ahamiyatga ega. Milliy korpus foydalanuvchilari turli sohadagi jurnalistlar, adabiyotshunoslar, tarixchilar, gumanitar ta'lim sohasi vakillaridir. Aytish mumkinki, ona tili va chet tilini o'qitishda, darslik va dasturlar tayyorlashda milliy korpusning ahamiyati juda katta. Binobarin, korpus lingvistikasining o'zbek tili ta'limining maxsus tarmog'i sifatida shakllanishi va yanada takomillashtirilishi o'zbek tilshunos-olimlari uchun keng ko'lamli eksperimental materiallardan foydalanish, zarur lingvistik ma'lumotlarni topish va ularga tegishli ishlov berish imkoniyatini yaratadi.

Milliy xususiyatga ega o'zbek tilining "shaxsiy" lingvistik korpusini tuzish muammosi hozirda O'zbekistonning bir qancha ilmiy va amaliy markazlarida muhokama qilinmoqda. Ammo korpusni tuzish markazlashtirilgan holda amalga oshirilsa quyidagi turdagi korpuslarni yaratish mumkin bo'ladi.

1) o'zbek tilidagi ommaviy axborot vositalarining zamonaviy (yoki davriy) matnlarining alohida korpusi (gazeta, jurnal sahifalarida);

- 2) o‘zbek so‘zlashuv tilining yozma (matn) alohida;
- 3) o‘zbek tilining multimedia korpusi (korpusning asosini matnlarning video va audio yozuvlari tashkil etadi);
- 4) o‘zbek tiliga parallel bo‘lgan alohida tillar korpusi (turkiy va hind-evropa tillari, parallel tillar korpusi);
- 5) o‘zbek tilining she’riy matnlarining alohida korpusi;
- 6) o‘zbek tilining o‘quv korpusi.

Xulosa o‘rnida shuni ta’kidlash joizki, korpus lingvistikasi tilshunosligimizda yangi yo‘nalish sifatida o‘z yechimini kutayotgan masalalarga to‘la. Soha rivojlanishi uchun tilshunosligimizda innovatsion texnologiyalar va nazariy tilshunoslik o‘rtasida uzluksiz aloqani o‘rnatish lozim.

Foydalanilgan adabiyotlar:

1. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
2. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (JFLTAL). 2019;6(1-2019):131-7.
3. Shaykhislamov, N. (2020). Cognitive Linguistics, the Symbolic and Interactive Functions of Language. Образование и наука в XXI веке, 1(6), 390-393.
4. A. Po‘latov, S. Muxamedova “Kompyuter lingvistikasi”. – Toshkent, 2007.
5. A. Hurnonov “Структур тилшунослик: илдизлари ва йўналишлари”. – Тошкент, 2008.
6. A.Rahimov “Kompyuter lingvistikasi asoslari” O‘quv qo‘llanma.-Toshkent, 2011.

О НЕОБХОДИМОСТИ СОЗДАНИЯ НАЦИОНАЛЬНОГО КОРПУСА УЗБЕКСКОГО ЯЗЫКА

Хидирова Гулнора Нарзиевна

Кандидат филологических наук, и.о.доцента

Ташкентский государственный экономический университет

Annotation: The relevance of the topic is due to new government decisions on the development of the Uzbek language, increasing its role and status as the state language. In the adopted “Concept for the development of the Uzbek language”, the creation of the National Corpus of the Uzbek language is one of the priority tasks. The beginning of the 21st century is marked by a powerful breakthrough of information technologies in the field of philological sciences: it became possible to teach the language through the so-called large arrays of texts. The article highlights some aspects of the purpose and role of the national corpus in the light of current trends in linguistics. The national corpus represents a given language at the present or a certain stage of its existence and in all the variety of genres, styles, territorial and social variants. The areas of application of the language corpus are very different: in the study of the grammar of a language, in linguodidactics (when teaching a language), in lexicography, in translation studies.

Keywords: corpus linguistics, electronic corpus, text, linguistic research, grammar, teaching, dictionaries, translation studies.

Аннотация: Актуальность темы обусловлена новыми правительственными решениями по развитию узбекского языка, повышению его роли и статуса как государственного языка. В принятой «Концепции развития узбекского языка» создание Национального корпуса узбекского языка является одной из приоритетных задач. Начало XXI века знаменуется мощным прорывом информационных технологий в сфере филологических наук: стало возможным обучение языку посредством так называемых больших массивов текстов. В статье освещены некоторые аспекты назначения и роли национального корпуса в свете актуальных тенденций лингвистики. Национальный корпус представляет данный язык на современном или определенном этапе его существования и во всём многообразии жанров, стилей, территориальных и социальных вариантов. Сферы применения языкового корпуса весьма различны: при изучении грамматики языка, при обучении языку, в лексикографии, в переводоведении. Посредством компьютерных технологий корпус превратился из метода работы с языковым материалом в эффективный инструмент исследования.

Ключевые слова: корпусная лингвистика, электронный корпус, текст, лингвистические исследования, грамматика, обучение, словари, переводоведение.

С обретением Независимости в Узбекистане были приняты ответственные решения по развитию узбекского языка, повышению его роли и статуса как государственного языка. В частности, одним из последних правительственных постановлений явилось утверждение Концепции развития узбекского языка и совершенствования языковой политики в 2020 – 2030 годах. “Сегодня наша масштабная работа по построению Нового Узбекистана

вступает в решающий этап. На этом пути родной язык как символ национального самосознания и гордости, безусловно, служит для нас источником огромной силы и вдохновения», - отмечает наш Президент Шавкат Миромонович Мирзиёев [Из речи Президента. 2021].

В Концепции определены такие приоритетные задачи как:

- открытие центров узбекского языка в зарубежных вузах;
- открытие новых направлений “Компьютерная лингвистика” и “Прикладная филология” в вузах филологического профиля;
- создание национального корпуса узбекского языка в электронном виде, включающего всю научную, теоретическую и практическую информацию об узбекском языке;
- создание компьютерной программы для перевода с узбекского языка на ведущие мировые языки и с иностранных языков на узбекский язык [3].

И в этой связи мы хотели осветить некоторые аспекты назначения и роли национального корпуса в свете актуальных тенденций лингвистики. Нужно отметить, что начало XXI века знаменуется мощным прорывом информационных технологий в сфере филологических наук: стало возможным обучение языку посредством так называемых больших массивов текстов. Со 2-ой половины XX века компьютерные технологии способствовали развитию практики создания корпусов, в связи с чем 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 доктор филологических наук Владимир Плунгян 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» [Плунгян Владимир. 2013].

Причиной стремительного развития корпусной лингвистики послужили формирующиеся современные интеллектуальные программные системы, предназначенные для обработки текстов на естественном языке. В то же время, эти системы требуют большой экспериментальной лингвистической базы. Эти два фактора – технические возможности и массивы текстов, на наш взгляд, и обусловили растущий спрос на корпусные данные. В настоящее время все современные лингвистические исследователи и составители словарей все больше используют объёмные корпусы текстов.

В настоящее время корпусы созданы для многих языков мира, в частности, для венгерского, испанского, датского, итальянского, словенского, польского, чешского, эстонского, хорватского, словацкого, японского, русского. Национальный корпус русского языка (НКРЯ) охватывает почти 4.5 млн русских текстов общей длиной более 1,5 млрд слов. В настоящее время в состав НКРЯ вошли 38 корпусов: Основной корпус, Газетный корпус, Региональный корпус, Поэтический корпус, Параллельный корпус, Синтаксический корпус, Диалектный корпус, Обучающий корпус, Устный корпус, Акцентологический корпус, Мультимедийный корпус, Мультимедийный параллельный корпус, Древнерусский корпус, Церковнославянский корпус, Корпус берестяных грамот и Старорусский корпус [Национальный корпус русского языка. 2003-2022.].

Таким образом, корпус - это база данных (информационно-справочная система), основанная на собрании текстов на некотором языке в электронной форме. Национальный корпус представляет данный язык на современном или определенном этапе его существования и во всём многообразии жанров, стилей, территориальных и социальных вариантов и т. п. Национальный корпус создается лингвистами для научных исследований и обучения языку.

В этой связи, считаем нужным отметить, что в Узбекистане в 2021 году был запущен проект создания Электронного корпуса узбекского языка под руководством профессора Национального университета - Абдурахмановой Н.З. Корпус узбекского языка представляет собой комплекс лингвистических словарей, веб-страниц, морфологической базы данных узбекского языка, учебной и научной литературы, официальных и художественных текстов различных жанров. Корпус содержит программное обеспечение с возможностью управления и использования данных. В перспективе корпус будет включать подкорпуса: Параллельный корпус, Авторский корпус, Образовательный корпус. К различным лингвистическим ресурсам относятся тезаурус узбекского языка, словарь англо-узбекских глагольных словосочетаний и другие лингвистические словари [8].

Итак, ключевым элементом корпусной лингвистики является языковой корпус (ЯК), в котором собраны и хранятся в электронном формате тексты разных жанров и стилей. «Его основное назначение (ЯК) состоит в обеспечении получения достоверной информации об употреблении слова и нахождении лексических единиц и грамматических конструкций благодаря лингвистической разметке» [2]. Джон Синклер в своем определении понятию корпуса, подчеркивал, что корпус «...функционирует как источник данных для лингвистических исследований» [7].

Структура корпуса напрямую связана с ее функцией и применением. Это означает, что для изучения и анализа определенного языкового уровня или категории (лексики, синтаксиса, стиля и т.п.) нужно собрать по возможности полную коллекцию текстов, касающихся этого уровня. В этой связи корпус объективно и оперативно подвергает анализу язык на так называемых «живых» примерах, т.е. «при работе корпусной лингвистики используются не искусственно созданные тексты, а примеры живого использования языка» [2].

Также отметим, что корпусная лингвистика включает не только исследование корпусов текстов, но также и корпусов словарей как регистраторов объемного словарного материала.

В настоящее время также активно рассматриваются методические аспекты использования информационной базы корпусных интернет-технологий в преподавании языков.

На сегодняшний день в мире функционируют множество одноязычных и многоязычных корпусов: Брауновский корпус, Британский национальный корпус, Национальный корпус американского английского, Национальный корпус русского языка, Википедия, Проект Татозба, Открытый корпус русского языка и мн. др. «Представительные корпуса существуют или разрабатываются для немецкого, польского, чешского, словенского, финского, новогреческого, армянского, китайского, японского, болгарского и других языков» [9].

Корпус, содержащий текстовую информацию на нескольких языках, именуется параллельным корпусом. К примеру, в Национальный корпус русского языка, в соответствии с мировым опытом построения параллельных корпусов, теперь входит многоязычный корпус, задействовавший более 20 языков [6].

Теоретические и практические исследования по созданию электронного корпуса и подкорпусов узбекского языка также проводятся в ряде вузов Узбекистана. В 2020-2021 годах в Ташкентском государственном университете узбекского языка и литературы осуществлялся проект по созданию учебного корпуса узбекского языка.

В 2021 году под руководством профессора Национального университета Узбекистана Н.З. Абдурахмоновой разработан проект по созданию Электронного корпуса узбекского языка. Современное состояние электронного корпуса узбекского языка представляет собой комплекс лингвистических словарей, веб-страниц, морфологической базы данных узбекского языка, учебной и научной литературы, официальных и художественных текстов различных жанров. Корпус содержит набор устных и письменных текстов, программное обеспечение с возможностью управления и использования данных. Планируется включить параллельный двуязычный (с узбекским) корпус для национального корпуса узбекского языка: английский, русский, турецкий и корейский.

Под руководством профессора Самаркандского государственного университета Суюна Каримова проводится практическое исследование «Проектирование национального корпуса узбекского языка и разработка программного комплекса». Основным результатом исследования станет программный комплекс для создаваемого Национального корпуса узбекского языка. В 2022 году успешно протестировано программное обеспечение корпуса, созданного на основе электронных текстов эпоса «Алпомиш».

Сферы применения языкового корпуса весьма различны, во-первых, при изучении грамматики языка корпусы дают возможность наблюдать за изменениями сочетаемости грамматических конструкций, появлением неологизмов, утратой номинации и мн. др.

Во-вторых, в лингводидактике корпусы представляются наиболее уместными, актуальными источниками языкового материала, т.к. полученная информация начинает определять содержание обучения. Мировая практика применения корпусной лингвистики в обучении свидетельствует об эффективности корпусных методов. Применение лингвистических корпусов при обучении языку значительно облегчают образовательный процесс. Они позволяют оценить употребление различных слов и выражений непосредственно носителями изучаемого живого языка. Онлайн-ресурсы лингвистических корпусов – удобный материал для использования при обучении языку в образовательном процессе. Учебные лингвистические корпуса активно используются преподавателями в методике обучения языку, а также студентами - в самостоятельных и исследовательских работах.

В-третьих, в лексикографии по принципу корпуса создаются бумажные и онлайн-словари, которые особенно ценны для переводчиков и преподавателей иностранных языков.

И, в-четвертых, в переводоведении, при обучении переводу, параллельные корпуса выявляют закономерности в тексте оригинала и текста перевода, так как одинаковая коммуникативная направленность в двух разноязычных текстах является объектом сопоставительных корпусов.

Объём и проработанность корпусов в составе Национального представляет широкие возможности для проведения достаточно сложных сопоставительных и типологических исследований, как на лексическом, так и на грамматическом уровнях. Создание Национального корпуса активизирует разработки на узбекском языке приложений к программным продуктам, что обеспечит значительный потенциал нашему языку в мировом информационном пространстве.

Таким образом, корпусная лингвистика – область лингвистики, в которой создаются, совершенствуются корпуса текстов, а также применяются в качестве инструмента лингвистического исследования.

Представительный массив языковых данных за определенный период позволяет исследовать динамику изменения лексического состава языка, анализировать лексико-грамматические характеристики разных жанров и у разных авторов.

Корпусы также служат инструментом в подготовке разнообразных исторических и современных словарей.

Данные корпусов используются прикладными лингвистами (преподавателями, переводчиками) для построения грамматик, в обучении языку и для решения своих профессиональных задач. Посредством компьютерных технологий корпус превратился из метода работы с языковым материалом в эффективный инструмент исследования. Возможности современных компьютерных технологий обусловили активное появление разнообразных корпусных проектов в мире, разнящихся масштабами, языками и целями. Во всем мире интерес к корпусной лингвистике только возрастает, поскольку данные корпуса – это не искусственно созданные тексты, а непосредственные примеры живого языка. Стало возможным проведение лингвистических исследований на большом речевом материале. Разработка правил организации текстов в корпус и методика их анализа, по сути, означало оформление корпусной лингвистики как самостоятельной лингвистической дисциплины со своими инструментами и методами исследования. Можно утверждать, что корпусная лингвистика формирует будущее лингвистики.

Использованная литература:

1. Из речи Президента, посвященной празднованию узбекского языка. 20 октября 2021 года.
2. Иргизова К. В. Корпусная лингвистика в отечественном и зарубежном языкознании на современном этапе. <https://cyberleninka.ru/article/n/...>
3. Концепция развития узбекского языка и совершенствования языковой политики в 2020 — 2030 годах. Указ Президента Республики Узбекистан от 20 октября 2020 года № УП-6084.
4. Национальный корпус русского языка. 2003-2022. Доступен по адресу: ruscorpora.ru
5. Плунгян Владимир. Корпусная лингвистика. 14.01.2013. <https://postnauka.ru/video/7783>.
6. Сичинава Д. В. Параллельные тексты в составе Национального корпуса русского языка: новые направления развития и результаты.- Стр.194.; Стр. 196.
7. Словарь английского языка Collins Cobuild. <https://www.livelib.ru/author/629033/top-dzhon-sinkler>
8. Узбекский электронный корпус. <https://uzbekcorpus.uz/rusVer>

KORPUS LINGVISTIKASI VA DUNYO TILSHUNOSLIGIDA KORPUSGA OID NAZARIYA VA TADQIQOTLAR

Vasliddinova Kamola Qodirjon qizi

Phd student

Uzbekistan State World Languages University

Annotatsiya: Bu maqolada korpusning tilshunoslik sohasida dastlab paydo bo'lishi, uning qadimgi hamda zamonaviy turlari va korpus atamasiga dunyo tilshunoslari tomonidan bildirilgan fikrlar haqida batafsil ma'lumotlar berilgan. Xorij ta'lim tizimida til o'rganish jarayonlarida bevosita korpusdan foydalanish va uning samaradorlik jihatlari yoritib berilgan.

Kalit so'zlar: Korpus, kompyuter lingvistikasi, korpus nazariyasi, arxeotipik korpus, korpus etikasi.

Annotation: This article provides detailed information about the first appearance of the corpus in the field of linguistics, its ancient and modern types, and the opinions expressed by world linguists on the term of corpus. The direct use of the corpus in the language learning process in the foreign education system and its effectiveness aspects are highlighted.

Keywords: Corpus, computational linguistics, corpus theory, archetypical corpus, corpus ethics.

Аннотация: В данной статье представлена подробная информация о первом появлении корпуса в области языкознания, его древних и современных видах, а также о мнениях, высказанных мировыми лингвистами о термине корпус. Выделены непосредственное использование корпуса в процессе изучения языка в зарубежной системе образования и аспекты его эффективности.

Ключевые слова: Корпус, компьютерная лингвистика, корпусная теория, архетипический корпус, корпусная этика.

XX-XXI asrlarda jamiyat hayotini axborotlashtirish va avtomatlashtirish har bir sohaga bevosita kirib bordi va bu jarayon tilshunoslik sohasini ham chetlab o'tmadi, balki tilshunoslik sohasida yangi tadqiqotlar va izlanishlar uchun muhim turtki vazifasini bajardi. Natijada tilshunoslikning amaliy va zamonaviy turlari bo'yicha yangi yo'nalishlar, atamalar va tadqiqot obyektlari paydo bo'ldi. Masalan, mashina (avtomatik) tarjimasini, kompyuter lingvistikasi, korpus lingvistikasi va bevosita bu yo'nalishlarda minglab termin va yangi so'zlar tilshunoslikka kirib keldi. Bir necha o'n yillardan beri korpus lingvistikasi turli dunyo olimlari tomonidan yangi tadqiqotlar markazidan oliy o'rin egallagani tilshunoslar uchun sir emas. Chunki korpus nafaqat har bir tilning milliylikni saqlab qolishda, uning bevosita tabiiylikni asrab keyingi avlodlarga yetkazib berishda, balki ana shu tilda yaratilgan milliy, madaniy va ilmiy boyliklarni jahon hamjamiyatiga sof holda, bevosita taqdim etishda xizmat ko'rsatuvchi elektron xazina, ya'ni elektron manbadir. E'tibor qaratsak, butun dunyoda rivojlangan mamlakatlarning tili dunyo sohasida keng ommalashib ulgurgani va bevosita turli rivojlangan sohalarda yangi paydo bo'layotgan terminlar ana shu tillardagi so'zlar bilan nomlanib, hattoki davlatlarning milliy

tillariga o'zlashtirilayotgani ana shu tillarning allaqachon avtomatlashtirib, turli millat vakillari foydalana olishi uchun qulaylashtirib quyilgani bilan izohlash mumkin. Masalan: biznes, trening, hotel, park, fast food, kafe, restoran, saboy, integrallashmoq, publika, stil, okey va boshqa o'nlab yangi so'zlar allaqachon o'zbek tili iste'molida faol hisoblanib, hattoki ommaviy axborot vositalarida ham muloqot jarayonida bevosita va bilvosita foydalaniladi. Bundan ma'lum bo'ladiki, tilning kompyuterlashtirilishi va til boyliklarining korpusga jamlanishi har qanday davlat rivojlanish bosqichlarida, o'zining ilmiy yutuqlari bilan yanada ildam qadam bosishi uchun ulkan yordamchi bo'la oladi.

Davlatimizda ham bugungi kunda amaliy tilshunosligimiz oldida turgan eng muhim masalalardan biri etib – o'zbek tilining milliy korpusini yaratish siyosat darajasiga ko'tarildi. Xususan, o'zbek tilini jamiyat hayotida hamda xalqaro miqyosda obro'-e'tiborini oshirish maqsadida, o'zbek tiliga oid barcha ilmiy, nazariy va amaliy ma'lumotlarni o'zida jamlagan elektron ko'rinishdagi o'zbek tili milliy korpusini yaratish, o'zbek tilini Internet jahon axborot tarmog'ida ommalashtirish, unda munosib o'rin egallashini ta'minlash, dasturiy mahsulotlarning o'zbekcha ilovalarini yaratish, o'zbek tilini o'rgatuvchi kompyuter dasturlarini keng miqyosda amaliyotga tatbiq qilish, o'zbek tilidagi matnlar tahririga mo'ljallangan kompyuter dasturlarini yaratish ishlari tilshunoslik oldida turgan muhim vazifalar etib belgilab olindi. [O'zbekiston Respublikasi Prezidenti Sh.Mirziyoyev, 2019; PF-5850]. Farmonning amaliy ijrosini ta'minlash uchun o'zbek tilshunos olim va olimalaridan B.Mengliyev, Sh.Shahobiddinova, Z.Xolmanova, S.Karimov, N.Abduraxmonova, L.Raupova, Sh.Hamroyeva, M.Abjalova, G.Toirova, G.Ikromova, J.Djumabayeva, G.Ergasheva, A.Eshmo'minovlar ilmiy tadqiqot ishlarini olib borishmoqda.

Korpus lingvistikasi bugungi kunda zamonaviy tilshunoslikning eng tez rivojlanib borayotgan sohasi bo'lib, korpus tushunchasiga olimlar tomonidan turli baxsli izohlar berilgan. Ingliz korpus tilshunosi Leech korpus lingvistikasiga quyidagicha ta'rif beradi: "Korpus lingvistikasi - kompyuter korpus lingvistikasi bo'lib, tilshunoslikda yangi falsafiy yo'nalish hisoblanadi" [Leech G.,1992: 106]. Lekin McEnery va Taylor korpus lingvistikasini tilshunoslikdagi yangi elektron qidiruv metodi sifatida izohlaydi va amaliy jarayonda Leechning fikrlari bilan deyarli kata farqlarni sezmagani ekanligini e'tirof etadi [cf. McEnery et al.2006:7f]. Sandra Kubler va Heike Zinsmeister "Corpus Linguistics and Linguistically Annotated Corpora" asarida korpus lingvistikasiga ham nazariya ham ma'lumotlar qidirish metodi sifatida izoh beradi va qay uslubda yondashish korpus tilshunosining ixtiyori ekanligini ta'kidlaydi [Hans Lindquist, 2009]. Biber va Togneni Bonelli esa zamonaviy tilshunoslikning aynan bu yangi yo'nalishi elektron korpusga asoslangan tilshunoslikning yangi yo'nalishi sifatida baholaydi va yuqoridagi fikrlarni tasdiqlaydi [Biber et al., 1998; Tognini-Bonelli, 2001]. Bugungi amaliy tajribalar shuni isbotlamodaki, korpus turli soha olimlari uchun ilmiy va nazariy fikrlarini kengroq ifodalab berishda va til o'rganish jarayonlarini osonlashtirishda yordam bera oluvchi elektron manbadir. Yanada soddaroq ifodalasak, *Korpus lingvistikasi* - bu korpuslarda saqlanadigan "real life"-kundalik hayotda insonlar tomonidan tabiiy jarayonda yozib olingan kommunikativ yoki chop etilgan publitsistik, ilmiy yoki siyosiy mantlar to'plamidir va lingvistik tadqiqotlar uchun yaratilgan kompyuter ma'lumotlar bazalariga asoslangan tilni o'rganish metodidir.

Korpusga asoslangan metodlar garchi 1960-yillarda paydo bo'lgan bo'lsa ham yangi yo'nalish sifatida 1980-yillardan tilshunoslikda paydo bo'ldi va korpus turli til tarmoqlarida yangi nazariya va yo'nalishlarni rivojlantirdi. Garchi XX asrdan boshlab korpus tushunchasi

tilshunoslikda paydo bo'lgan bo'lsada qadimdan korpusning arxeologik ko'rinishi mavjud bo'lgan. Bunga yorqin misol kutubxonalar bo'lib, ma'lumotlar qog'oz shaklda asrlar osha saqlab kelingan. XIII asrda Christian Bible tomonidan **Archeotypical corpus** (arxeotipik korpus)ga asos solingan bo'lib, yeg'ilgan matlar *word indexing*(so'zni indekslash) yoki *concordancing*(so'zlarning o'zaro munosabati va bog'lanishi) orqali tartib bilan joylashtirilgan[<https://www.jw.org/en/library/bible/nwt/bible-words-index/>]. Lekin inson hayotida avtomatlashtirishning keng yoyilishi bu usulning elektron variantini yanada mukammallashtirilgan ko'rinishini (*corpus based*)amaliyotga tadbqiq etdi. Hattoki tilshunos Charles Fillmore tomonidan korpusga asoslangan tilshunoslik yo'nalishiga (**computer-aided armchair linguistics**) ya'ni kompyuter yordamida kresloga suyanib olgan tilshunoslik deya ta'rif berildi[CH Fillmore, 1992]. Chunki bu yangi metod Christian Biblesining word indexing va concordancingga asoslangan an'anaviy arxeotipik korpusidan soddaligi, qulayligi va ishonchliligi bilan nafaqat korpus tilshunosi, balki har qanday foydalanuvchi yoki til o'rganuvchi uchun samarali manbadir.

Ma'lumki korpusga asoslangan dastlabki tadqiqotlar ingliz tilida olib borildi va dunyo tilshunoslari tomonidan tadqiqotlarning asosiy markaziga aylandi. 1999-yilda Germaniyaning Freiburg universitetida tashkil etilgan **ICAME(the International Computer Archive of Modern/Medieval English)**ning har yili tashkil etiladigan forumining shiori **Corpus linguistics and linguistic theory**(Korpus lingvistikasi va lingvistik nazariya) deb nomlandi. Tadqiqotlar samarasi o'laroq birinchi zamonaviy korpus **the Brown University Standard Corpus of Present-day American English** 1960-yilda Brown universitetida yaratilgan va taxminan bir millionta so'zdan iborat bo'lgan matnlarni na'muna sifatida kiritgan [Francis and Kuchera, 1967]. Na'munaga asoslangan holda turli davlat tilshunoslari tomonidan o'z milliy tillarining korpusini shakllantirish XX asr oxirlaridan ommalashib bordi. Masalan LOB korpus(Lancaster-Oslo/Bergen), BNC(British National Corpus), ICLE(International Corpus Of Learner English), COLT(Corpus Of London Teenage Speech), ICE-GB(The International Corpus Of English British Component), ONZE korpus(The Origins Of New Zealand English), COCA(The Corpus Of Contemporary American English), The BoE korpus(The Bank Of English), the LCMC(the Lancaster Corpus Of Mandarin Chinese)[Tony McEnery and Andrew Hardie, 2012] va boshqalar.

Har qanday korpusning shakllanishi mobaynida to'plangan matnlarning etikasi(corpus ethics), foydalanish chastotasi(use of frequency) va olinayotgan materialning qonuniy masalalari(legal issues)ga e'tibor qaratish lozim. Har qanday mamlakat o'z milliyligi va ma'naviy – madaniy qadriyatlariga ega bo'lib axloqiy me'yor darajasi dini yoki madaniyati jihatidan bir biridan qisman yoki tubdan farqlanadi. Korpusga matnlar jamlash jarayonida tilning tabiiyligini(NLP) saqlash ustuvor vazifa hisoblanib, jarayon davomida kollektorlar tomonidan turli elektron va yozma manbalarga murojat qilish mumkin. Ayniqsa XXI asr korpus tilshunoslari uchun axborot qidiruv(search of information) imkoniyatlarini ham oson, ham ba'zi o'rinlarda qiyinlashtirdi. Chunki **The World Wide Web** ning global miqyosda ommalashib borishi har qanday foydalanuvchi(**user**) uchun ishonchli va soxta(**fake**) ma'lumotlarning iste'mol jarayonida qo'shib ketishiga sabab bo'ldi desak mubolag'a bo'lmaydi. Korpus tilshunosi tomonidan materiallarning sof, qonuniy, tabiiy va etik jihatdan mos variantlarini ajratib olish bir muncha qiyinlashdi.

Korpus sohasida amalga oshirilayotgan izlanishlar natijasida dunyo ta'lim tizimi rivojlanish bosqichida yanada ildamladi. Chunki rivojlangan mamlakatlar ta'lim tizimida

korpusdan dars jarayonlarida bevosita ta'lim sifatini oshirishda va dars jarayonlarida noan'anaviy (*indirect*) metodni keng targ'ib etishda foydalanish an'anaviy usulga aylandi. Bunga yorqin misol sifatida o'qituvchi til o'rgatish jarayonida bevosita so'z tarjimasi yoki ma'lum bir xorijiy matn bilan ishlashda to'g'ridan to'g'ri korpusga murojat etishi til o'rganish jarayonlarini yanada oson va soddalashtirishini tajribalar allaqachon isbotladi. Dunyo tajribasida jami o'n ikkita metodga asoslangan korpus yordamida til o'rgatish, o'rganish yoki tajriba uchun foydalanishga mo'ljallangan kitob holiga keltirilgan qo'llanma bo'lib [Ken Hyland, 2013], nafaqat o'rta, balki oliy ta'lim dasturlarida ham bugungi kunda dunyo reytingidan yuqori o'rin egallagan mamlakat universitetlari tomonidan ham keng foydalanib kelinmoqda. Hyland tomonidan yozilgan "Corpora and innovation in English language education" kitobida tajriba dasturlarga batafsil to'xtalib o'tilgan bo'lib, o'zbek tili milliy korpusini yanada mukammal va foydalanish uchun qulay holga keltirish uchun korpus sohasida qo'lga kiritilgan xorij ilmiy yutuqlarini batafsil o'rganib chiqish muhim vazifalardan biri sanaladi.

References:

1. cf. McEnery et al.2006:7f
2. Francis and Kuchera, 1967
3. Hans Lindquist, Corpus Linguistics and the Description of English. Edinburgh University Press, 2009
4. <https://www.jw.org/en/library/bible/nwt/bible-words-index/>
5. Ken Hyland. Corpora and innovation in English language education, 2013
6. 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
7. Leech, G. Corpora and theories of linguistic performance. 1992.-p.106
8. 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.
9. N. Z. Abdurakhmonova, A. S. Ismailov and D. Mengliev, «Developing NLP Tool for Linguistic Analysis of Turkic Languages,» 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation, 2022, pp. 1790-1793, doi: 10.1109/SIBIRCON56155.2022.10017049.
10. O'zbekiston Respublikasi Prezidenti Sh.Mirziyoyev, 2019; PF-5850
11. Tony McEnery and Andrew Hardie, Corpus Linguistics: methods, theory and practice. Cambridge University Press, 2012

KORPUS–TABIY TILLARNING MILLIY XUSUSIYATINI SAQLAB QOLISHDA MUHIM INSTRUMENT

Yulbarsov Ochilbek Obidjon o'g'li

O'zbekiston Milliy universiteti O'zbek filologiyasi
fakulteti o'qituvchisi, mustaqil izlanuvchi

Rustamova Zulfiya Xolmirzayevna

Nizomiy nomidagi TDPU o'qituvchisi,
mustaqil izlanuvchi

Annotatsiya: Maqolada Korpus lingvistikasi sohasining predmeti Korpus tushunchasining tilshunoslikdagi ahamiyati hamda uning funksional vazifalariga to'xtalib o'tilgan. Tabiiy tillarning milliy xususiyatini saqlab qolishdagi o'rni yoritib berilgan. Korpusning paydo bo'lishi, olimlarning qarashlariga qisman to'xtalib o'tilgan.

Kalit so'zlar: korpus, tilshunoslik, uzbekcorpus milliy til, Kadling, tabiiy til, Douglas Biber, milliy xususiyat, lingvistik texnologiya.

Abstract: The article focuses on the subject of Corpus Linguistics, the importance of the concept of Corpus in linguistics and its functional tasks. The role of natural languages in preserving the national character is highlighted. The emergence of the corpus, the views of scientists have been partially touched upon.

Keywords: corpus, linguistics, uzbekcorpus.uz, national language, Kadling, natural language, Douglas Bieber, national feature, linguistic technology.

Аннотация: Статья посвящена теме корпусной лингвистики, значению понятия Корпус в языкознании и его функциональным задачам. Выделена роль естественных языков в сохранении национального характера. Частично затронуты взгляды ученых на возникновение корпуса.

Ключевые слова: корпус, лингвистика, узбекский корпус, национальный язык, Кадлинг, естественный язык, Дуглас Бибер, национальный признак, лингвистическая технология.

Ma'lumki, Davlatimiz tomonidan bugungi kunda ilm-fan taraqqiyoti uchun dunyoning rivojlangan mamlakatlari bilan ochiq siyosat olib borilmoqda. Doimo manfaatli hamkorlikka intilib kelinmoqda. Bu esa fandagi yangi tushuncha va nazariyalarni kashf etilishiga hamda amaliyotga joriy etilishida katta hissa bo'lib xizmat qilmoqda. Yangi sohalarni kirib kelishi va rivojlanishi esa mamlakat va jamiyat taraqqiyoti uchun muhim omil bo'la oladi. Jumladan, Tilshunoslik sohasida ham hozirgi kundagi global vazifa – tilni tadqiq qilish va uni amaliyotga tatbiq etishda zamonaviy texnologiyalaridan maqsadli foydalanishdan iborat desak yangilishmagan bo'lamiz.

Tilshunoslikka kirib kelayotgan har bir yo'nalish yoki terminlarni o'z sohasi o'laroq tahlil qilish va tasnif etish, undan amaliy tarzda foydalanish, o'zbek tilini axborot-kommunikatsiya

texnologiyalari orqali qo'llash kabi ishlar til tadqiqining ustuvor maqsadlari safiga qo'shishni taqozo etadi.

Global dunyoning jadallik bilan o'zgarib borishi milliy tilning ijtimoiy-ommaviy ta'sirlardan munosib himoya qilish uchun uni elektron shakllarini yaratishni taqazo qilmoqda. Xususan, Korpus lingvistikasi sohasi tilning material sifatidagi mavqeyini saqlay oladi. O'zbek tilshunosligida nisbatan yangi soha hisoblanuvchi Kompyuter lingvistikasining ajralmas qismi hisoblanadigan Korpus lingvistikasi sohasi ham tilning amaliy – matn aspektidagi muhim instrumentlardan deyish mumkin.

XXI asrning global muammolaridan biri tabiiy tillarning milliy xususiyatini saqlab qolishdan iborat ekan dunyo tillarini elektron korpuslarini yaratish va rivojlantirishda NLP hamda til texnologiyalariga doir tadqiqotlarni izchil ravishda olib borish dolzarb vazifaga aylandi. [Abduraxmonova N. 2021:135] Demak, bugungi kunda Korpus tushunchasi o'z funksiyalarini til va Tilshunoslik sohasida amaliy tarzda taqdim etuvchi sifatida namoyon bo'lmoqda. Bu terminni tilimizda juda ko'p o'rinlarda ishlatib kelmoqdamiz.

Wikipedik ma'lumotga ko'ra, Ko'rpus (lotincha: Corpus – tana) -1) inshoot, mexanizm, asboblari, apparat, qurilmalar va ba'zi transport mashinalarining qismi; 2) yaxlit maydondagi bir necha binodan bittasi yoki yirik inshootning alohida qismi; 3) kegel (o'lchami) 10 punktga (3,76 mmga) teng bo'lgan bosma shrift; 4) biror maxsus ish bilan mashg'ul bo'lgan mutaxassislar hay'ati (mas, diplomatik korpus, ofitserlar korpusi va boshqalar). [7]

Korpus – bu matnlar majmuasi. Ya'ni ma'lum sohadagi matnni lingvistik nuqtayi nazardan tasniflovchi elektron to'plam. Bu matnlar turli jihatga ko'ra tasniflangan holatda bo'lishi ham mumkin. Ular kompyuterning ma'lumotlar bazasida saqlanadi. Matnlar og'zaki va yozma shaklda bo'ladi.

Ilk korpusning yaratilishi 1812-yilga borib taqaladi, bunda nemis olimi Kadling o'zining nemis so'zlaridagi undosh harflar distributsiyasini tahlil qilgan. Vaholangki, u davrda hali kompyuter terminining o'zi ham bo'lmagan. Keyinchalik zamonaviy ingliz tili korpusi namunalardan biri sifatida Broun korpusi 1960-1961 yillarda yaratildi va u ilk bor bosma holda chop etildi. Oradan bir yil o'tibgina ushbu korpus elektronlashtirildi. Ko'rinib turibdiki, ilk korpuslar kompyuter texnologiyalarisiz ham mavjud bo'lgan va izlanishlar olib borilgan. Keyinchalik fan rivoji qidiruv metodi (konkordans)ni elektron formatdagi matnlarda o'tkazishni taqazo etdi va kompyuter lingvistikasi bilan sohalararo munosabatga ehtiyoj tug'ildi.

Shu bilan birga kompyuter tilshunosligi odatda, kompyuter vositalari (dasturlari, ma'lumotlarni tashkil qilish va qayta ishlash uchun kompyuter texnologiyalari)ni muayyan sharoitlarda, vaziyatlarda, muammoli sohalar va tillardagi modellarning ko'lamini nafaqat tilshunoslikda, balki boshqa fanlarga ham qo'llashni nazarda tutadi. [“Ўзбекистонда хорижий тиллар” илмий-методик электрон журнал/ №2-2019/. 38 б.]

Korpus tushunchasi hozirgi kundagi lingvistik tadqiqotlarda, xususan, amaliy tilshunoslikda faol qo'llana boshlaganini uchratishimiz mumkin. Ko'rishimiz mumkinki, bu atamaning tilda fan tarmog'iga aylanib borishi aynan tilning material sifatidagi yo'nalishlari uchun xususiy tahlillarga olib kelgan. Shuning uchun ham bugungi kunda Kompyuter lingvistikasining ajralmas tarmog'i sifatida Korpus lingvistikasi alohida – konkret nazariyalarga ega soha deyish mumkin. Ko'rpus lingvistikasini bir qancha funksiyalari mavjud bo'lib, ulardan eng muhimi matnli resurslarning kompyuter bazasini yaratishdan iborat. Matnlar bazasi esa har xil sohalar

uchun ma'lumotlarni tezkor qidirib topish hamda undan o'z tadqiqotlarida foydalanishda olimlarga juda katta yordam bermoqda.

Korpus lingvistikasining hozirda alohida soha yoki fan bo'lib shakllanishida olimlarning nazariy qarashlari katta hissa qo'shgan. Korpus lingvistikasi asosan Kompyuter lingvistikasi bilan birgalikda uning bir yo'nalishi deb baholangan bo'lsa, ayrim nazariy qarashlarda alohida funksiyaga ega fan sifatida o'rganiladi. Shuningdek Korpus lingvistikasi sohasida nazariy hamda amaliy tadqiqotlar olib borgan olimlar soha rivojiga salmoqli hissalarini qo'shganlar. Xususan, Kading, Fries va Traver, Bongers, Kennedy, Stefan Gries, L.N. Zasorin, V. Zaxarov, Dougles Biber kabi dunyo olimlari hamda O'zbekistonda Kompyuter lingvistikasi sohasi shakllanishiga ulkan hissa qo'shgan professor A. Po'latov, ilmiy tadqiqotlar olib borgan N. Yoqutova, M. Ayimbetov, S. Rizayev va S. Muhamedov va boshqalar ilmiy-amaliy ishlar olib borgan. [Abduraxmonova.N., 2021:284-, 285-b.] Bundan tashqari, N.Z.Abduraxmonova "Kompyuter lingvistikasi" nomli darsligida tabiiy tillarni qayta ishlashning lingvistik asoslarini yaratish hamda korpus lingvistikasiga oid nazariy va amaliy masalalarni o'zbek tili doirasida o'rganadi. Olimaning muallifligida uzbekcorpus.uz – o'zbek tilining elektron korpusi yaratildi. Bu korpusni ayni paytdagi holatida o'zbek tilida yaratilgan lug'atlar, web sahifalar, o'zbek tilining morfologik ma'lumotlar bazasi, o'quv adabiyotlar hamda turli janrdagi ilmiy, rasmiy va badiiy matnlar majmuasi tashkil topgan [6]. Nazariy hamda amaliy tadqiqotlar o'zbek tili korpusi uchun muhim qirralarini jadallik bilan kashf qilib bormoqda.

Korpus lingvistikasi haqida olimlarning nuqtayi nazarlari turlicha. Qayd etilishicha, u til modeli emas, muayyan darajada metodologik yondashuv sifatida qarash to'g'ri bo'ladi. Dougles Biber uni quyidagi sifatlarini sanab o'tadi:

- u tabiiy matndagi zarur birliklarni empirik tahlil qiladi;
- analiz uchun "Korpus" sifatida tabiiy matnlarning katta va tizimlashtirilgan jamlanmalarini birlashtiradi;
- analiz uchun kompyuterning ham avtomatik va interaktiv texnologiyalardan foydalanish imkonini beradi;
- u analitik texnologiyaning miqdor (statistik) va sifat xususiyatlarini o'z ichiga oladi [Abduraxmonova.N., 2021:284-b.]

Aytib o'tganimizdek Korpus tilning har qaysi sohadagi turli maqsadlari uchun faol xizmat qiladigan zamonaviy lingvistik texnologiya. Tilshunoslikda ham korpusni tilning funksional jihatdan tasniflovchi muhim "qurilma" deyishimiz mumkin. Ayniqsa lug'atlarda hamda leksikografik ma'lumotlarni shakllantirishda samarali instrument deb qarashimiz mumkin.

Yuqorida sanab o'tilgan olimlarning qarashlarini hamda Korpus lingvistikasining tilda dolzarblik tendensiyalarini o'rganar ekanmiz, bu omillar lingvistikaning boshqa sof elementlarini ham elektron korpusga kiritish, tahlil qilish zaruratini yuzaga keltirmoqda. Xususan, tilshunoslikdagi leksikografiya tushunchasi va uning amaliy ko'rinishi ham zamonaviy amaliy kompyuter leksikografiyasining rivojlanishi uchun muhim manba bo'la oladi. Lug'atning har qanday ko'rinishi hozirda elektron tarzga ko'chib borayotganligi undan foydalanishda qulayliklarni yaratib bermoqda. Bundan tashqari Muloqot jarayonida neologizmlarni Korpus lingvistikasi asosida tasniflash, instrument sifatida qarash korpusning ijtimoiy tilga moslashuvchanligiga misol bo'ladi.

Xulosa o'rnida aytish mumkinki, Korpus boshqa til vakillarini ma'lum tilni yaxshiroq tushunish, nazariy hamda amaliy aniq xulosaga kelishi uchun muhim instrument sifatida namoyon bo'ladi. O'zbek tilida ham biror so'zni konteks bilan namoyon bo'lishi yoki matn tarkibida kelishi til o'rganuvchilar uchun qulaylik keltirib chiqaradi. O'zbek tilshunosligi doirasida Korpus lingvistikasi sohasini rivojlantirilishi milliy tilimizning kelajagi uchun qo'yilgan muhim qadamlardan hisoblanadi.

Adabiyotlar ro'yxati:

1. Abduraxmonova N. O'zbek tili elektron korpusining kompyuter modellari (monografiya). / Toshkent: Muharrir, 2021
2. Н.Б. Атабозев. Корпус лингвистикасининг асосий хусусиятлари.
www.journal.fledu.uz/ “Ўзбекистонда хорижий тиллар” илмий-методик электрон журнал/ №2-2019
3. N. Z. Abduraxmonova. “Kompyuter lingvistikasi” Darslik. Toshkent – 2021
4. И. И. Саженин словарный корпус как элемент оптимизации исследовательского процесса
5. Н. Jamolxonov. Hozirgi o'zbek tili. Toshkent – 2015
6. www.uzbekcorpus.uz
7. <https://uz.wikipedia.org/wiki/Korpus>

SEMANTIK ANNOTATSIYALANGAN KORPUSLAR TAVSIFI

Ismadiyorov Zafar

O'zbekiston Milliy universiteti, Kompyuter lingvistikasi va amaliy
tilshunoslik mutaxassisligi 1-bosqich magistranti

Kamilov Temur

O'zDJTU VA ADU qo'shma ta'lim dasturi,
Kompyuter lingvistikasi yo'nalishi 2-bosqich talabasi

Annotasiya: Ushbu maqola zamonaviy korpus lingvistikasida semantik annotatsiyalarning tutgan o'rni hamda uning vazifalari haqida so'z boradi. Ushbu maqola mavzusini yoritishda xorij va milliy fondagi ilmiy adabiyotlar o'rganilib mavjud muammolar yuzasidan tegishli taklif va tavsiyalar keltiriladi.

Kalit so'zlar: Korpus lingvistikasi, semantik annotatsiya, tabiiy til, uslub.

Annotation. This article discusses the role of semantic annotations in modern corpus linguistics and its functions. In covering the topic of this article, the scientific literature of foreign and national background is studied and the relevant suggestions and recommendations are given regarding the existing problems.

Keywords: corpus linguistics, semantic annotation, natural language, style.

Аннотация: В данной статье рассматривается роль семантических аннотаций в современной корпусной лингвистике и их функции. При освещении темы данной статьи изучается научная литература зарубежного и отечественного происхождения, даются соответствующие предложения и рекомендации относительно существующих проблем.

Ключевые слова: корпусная лингвистика, семантическая аннотация, естественный язык, стиль.

Korpus lingvistikasi tilni o'rganishning metodologik yondashuvi bo'lib, u korpus deb nomlanuvchi katta matnlar to'plamini tahlil qilishni o'z ichiga oladi. Boshqa bir ta'riflarga ko'ra, korpus-yozma matnlar to'plami yoki og'zaki til yoki ikkalasining kombinatsiyasi bo'lishi mumkin. Korpus turli sohalar, janrlar va registrlardagi inson tomonidan yozilgan matnlar to'plamini o'z ichiga oladi, masalan: yangiliklar maqolalari, akademik maqolalar va adabiy asarlar. Har bir matn aniq nutq bog'lovchilari bilan izohlanadi, ular ikki yoki undan ortiq nutq birliklari o'rtasidagi munosabatni bildiruvchi so'z yoki iboralardir.

Korpus lingvistikasining asosiy yo'nalishi tildan foydalanishdagi qoliplarni, jumladan grammatik tuzilmalarni, ya'ni lug'at va nutqni tahlil qilishga qaratilgan. Ushbu tahlillarni amalga oshirishda olimlar ilmiy tilda *“semantik annotatsiya”* atamasini qo'llaydi. Semantik annotatsiya bu- tabiiy tildagi matnni ma'lumotlar bilan belgilash jarayonidir. Semantik izohli korpuslar tabiiy tilni qayta ishlash va hisoblash lingvistikasi tadqiqotlari uchun muhim manba hisoblanadi [Николаев И.С., Митренина О.В., Ландо Т.М. (ред.) 2016. - 320 с.]. Korpus lingvistikasidagi semantik annotatsiya korpus ichidagi turli elementlarga ma'no berishni o'z ichiga oladi. Bunga

alohida soʻzlar yoki iboralar uchun teg belgilash, nomlangan obyektlarni aniqlash va soʻzlar yoki iboralar oʻrtasidagi munosabatlarni oʻrnatish kiradi. [Abduraxmonova N. (avtoreferat. 2021. 72 b.)]

Semantik izohlash jarayoni, odatda, anʼanaviy usul yordamida amalga oshiriladi va tegishli maʼlumotni aniqlash va belgilash uchun korpus boʻylab ishlaydigan annotatorlar guruhini oʻz ichiga oladi.

Korpus lingvistikasida semantik annotatsiya qoʻllanilishiga koʻra nutq va muloqot jarayonini tahlil qilishdir. Semantik maʼlumot uchun korpusga izoh berish orqali tadqiqotchilar maʼno belgilarini aniqlashlari va tildan maʼlum xabarlar yoki gʻoyalarni yetkazish uchun qanday foydalanishni tahlil qilishlari mumkin. Bu korpusdagi asosiy mavzular yoki mavzularni aniqlash, shuningdek, metafora, yumor va boshqa ritorik vositalardan foydalanishni oʻrganishni oʻz ichiga olishi mumkin. Semantik annotatsiya bu tabiiy tildagi matn maʼlumotlar bilan belgilash jarayonidir. Semantik izohli korpuslar tabiiy tilni qayta ishlash va hisoblash lingvistikasi tadqiqotlari uchun muhim manba hisoblanadi. Bu korpuslar, odatda, katta hajmdagi matn, jumla yoki ibora izohlarini tushunishga hamda semantik teglar yoki mashinani oʻrganish modellariga matn zahiridagi maʼnoni tushunishga yordam beruvchi teglar jamlanmasini oʻz ichiga oladi [Abduraxmonova N. (avtoreferat. 2021. 72 b.).]

Tadqiqotchilar semantik annotatsiylangan korpuslarini quyidagi guruhlariga ajratib tahlil qiladi:

PropBank - bu Penn Treebank predikat semantikasi uchun izohlarni taqdim etadigan korpusga asoslangan resurs hisoblanadi. U 2005-yilda Marta Palmer va uning jamoasi tomonidan ishlab chiqilgan. Xususan, PropBankdagi har bir predikat va uning argumentlari yoki shu jarayonda ishtirok etuvchi tarkibiy qismni ifodalovchi semantik rollar toʻplami bilan bogʻlangan. PropBank turli xil tabiiy tillarni qayta ishlash (NLP) vazifalari, masalan, tahlil qilish, semantik rollarni belgilash va mashina oʻrganish modeliga asoslangan tarjima jarayonida keng qoʻllaniladi. U predikatning argument tuzilishi va hisoblash semantikasini oʻrganish uchun muhim manba sifatida baholanadi [<https://www.cs.rochester.edu/~gildea/palmer-propbank-cl.pdf>].

FrameNet - bu Charlz J. Fillmor va uning jamoasi tomonidan 1998-yilda ishlab chiqilgan korpusga asoslangan leksik resurs hisoblanadi. Resurs vazifasiga koʻra, leksik maʼnoning ramka shaklida ifodalanishini taʼminlaydi, shu jumladan, mazkur jarayonda ramka turli komponentlarni yoki konseptual tuzilmani ifodalaydi. Maʼlumot oʻrnida taʼkidlash joizki, FrameNet korpusida asosiy ramkalar va ular bilan bogʻliq semantik rollarni aniqlash uchun qoʻlda tahlil qilingan izohli jumalarning katta toʻplami mavjud. [<https://catalog.ldc.upenn.edu/LDC2019T05>]

Penn Discourse TreeBank (PDTB)- korpusga asoslangan maʼlumotlar toʻplami boʻlib, matndagi nutqiy munosabatlari yoki bogʻlovchi elementlar uchun izohlar beradi. U 2000-yillarning oʻrtalarida Pensilvaniya universitetida Rashmi Prasad va uning jamoasi tomonidan ishlab chiqilgan. PDTB tabiiy tilni qayta ishlash (NLP), hisoblash tilshunosligi va diskurs tahlili kabi turli sohalaridagi tadqiqotchilar uchun kerakli resursdir. Resurs nutqni tahlil qilish, savollarga javob berish va umumlashtirish uchun bir qator hisoblash modellarini ishlab chiqish uchun foydalaniladi. [<https://catalog.ldc.upenn.edu/LDC2019T05>]

CallHome Japanese corpus - yapon tilida soʻzlashuvchilar oʻrtasidagi telefon suhbatlarida qoʻllaniladigan umumiy leksik birliklar toʻplamidir. U lingvistik maʼlumotlar konsorsiumi tomonidan yaratilgan. Korpus taxminan 1200 ta suhbatdan iborat boʻlib, nutqning umumiy davomiyligi taxminan 70 soatni tashkil qiladi. Suhbatlar turli mavzularni, jumladan, shaxsiy munosabatlar, kundalik tartiblar va dolzarb voqealarni qamrab oladi. Shuningdek,

transkript va izohlarni o'z ichiga oladi, bu esa uni nutqni aniqlash, tilni modellashtirish va suhbat tahlili bo'yicha tadqiqotlar uchun qimmatli manbaga aylantiradi.
[<https://catalog.ldc.upenn.edu/LDC2019T05>]

SenSem project for Spanish- ispan tili uchun semantik tarmoq yaratishga qaratilgan tizim bo'lib, bu kompyuterlarga turli kontekslarda so'zlarning ma'nosini yaxshiroq tushunish imkonini beradi. Loyiha Madriddagi Complutense universiteti tadqiqot guruhi tomonidan amalga oshirilmoqda va Ispaniya hukumati tomonidan moliyalashtirilgan.[https://www.researchgate.net/publication/230876727_Spanish_Framenet_A_frame-semantic_analysis_of_the_Spanish_lexicon]

Mazkur yuqorida keltirilgan guruhlarining asosiy vazifasi: korpus lingvistikasida semantik annotatsiyani o'rganish, tildan foydalanishni tahlil qilish uchun yangi texnika va vositalarni ishlab chiqishga olib keldi va matnni umumlashtirish, savollarga javob berish va hissiyotlarni tahlil qilish kabi bir qator tabiiy tillarni qayta ishlash ilovalarining aniqligi va samaradorligini oshirishga imkon beradi.

Xulosa qilib aytganda, o'zbek korpus lingvistikasida semantik annotatsiyaning asosiy vazifalarini aniqlash yuzasidan tadqiqotlarni yanada kuchaytirish lozim. Chunki, ushbu tushunchaning mohiyatini o'zbek tilidagi adabiyotlar yordamida tasniflashda manbalar, tadqiqot ishlari yetarli emasligiga guvoh bo'ldik.

Foydalanilgan adabiyotlar:

1. Николаев И.С., Митренина О.В., Ландо Т.М. (ред.) Прикладная и компьютерная лингвистика, Ленанд, 2016. - 320 с. - ISBN 978-5 9710-3472-8.
2. Abduraxmonova N. O'zbek tili elektron korpusining kompyuter modellari (avtoreferat). Toshkent: "AKTIV PRINT", 2021. 72 b.
3. N. Z. Abdurakhmonova, A. S. Ismailov and D. Mengliev, «Developing NLP Tool for Linguistic Analysis of Turkic Languages,» 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation, 2022, pp. 1790-1793, doi: 10.1109/SIBIRCON56155.2022.10017049.
4. 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.
5. 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.

KOMMUNIKATIV TIL IMKONIYATLARINI RIVOJLANTIRISHDA KORPUS TEXNOLOGIYALARIDAN FOYDALANISH

Sadikova Munisa Kuanish qizi,

Chirchiq davlat pedagogika universiteti magistranti

E-mail: sadikovamunisa522@gmail.com

Annotatsiya: Ushbu maqolada kommunikativ til imkoniyatlarini, ya'ni muloqot kompetentligini rivojlantirishda korpus texnologiyalaridan foydalanishning ahamiyatliligi yoritib berilgan. Kompyuter nutq faoliyatining barcha turlari ustida ishlash uchun foydalanish imkonini beradi, nafaqat o'qish va yozish ko'nikmalarini, balki tinglash va nutq ko'nikmalarini (shu jumladan talaffuz va intonatsiya kabi jihatlarni) rivojlantirish ham ahamiyatlidir.

Kalit so'zlar: o'zbek tili, kommunikativ, korpus, nutq ko'nikmalari, integratsiya.

Аннотация: В данной статье подчеркивается важность использования корпусных технологий в развитии коммуникативных языковых возможностей, то есть коммуникативной компетенции. Компьютер можно использовать для работы над всеми видами речевой деятельности, важно развивать не только навыки чтения и письма, но и навыки аудирования и говорения (включая такие аспекты, как произношение и интонация).

Ключевые слова: узбекский язык, коммуникативный, корпус, речевые навыки, интеграция.

Annotation: This article highlights the importance of using corpus technologies in the development of communicative language capabilities, that is, communication competence. The computer can be used to work on all kinds of speech activities, it is important to develop not only reading and writing skills, but also listening and speaking skills (including aspects such as pronunciation and intonation).

Keywords: uzbek language, communicative, corpus, speech skills, integration.

So'nggi 20-25 yil ichida dunyo barchamizni, barcha aloqa usullarini qamrab olgan eng katta inqilobni boshdan kechirmoqda. Ushbu inqilob yangi texnologiyalarni faol joriy etish bilan bog'liq. Respublikamiz Prezidenti Sh.Mirziyoyevning 2020-yil 20-oktabrdagi "Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risida" gi farmoni bilan tasdiqlangan "2020-2030-yillarda o'zbek tilini rivojlantirish va til siyosatini takomillashtirish konsepsiyasi" da belgilangan davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga faol integratsiyalashuvini ta'minlash ustuvor yo'nalishidagi o'zbek tiliga oid barcha ilmiy, nazariy va amaliy ma'lumotlarni o'zida jamlagan elektron ko'rinishdagi til imkoniyatlarini rivojlantirishda korpus texnologiyalaridan foydalanish muhim ahamiyatga ega. O'zbek tili dunyoning rivojlangan tillari qatorida turadi. Shu bilan birga kompetensiyalardan foydalanish til imkoniyatlarini rivojlantirishda muhim ahamiyatga ega. XX asrda o'qish va yozishni bilmaydigan odamni savodsiz deyishsa, XXI asrda o'rgangan bilimlaridan amaliyotda foydalana olmaydigan odamni savodsiz deyishadi.

Korpus –bu bir necha yoki muayyan til matnlarining yig‘indisiga asoslangan elektron shaklda to‘plangan ma’lumot (so‘rovnoma)lar tizimidir. 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’minot hisoblanadi.

Misol sifatida bitta mavzu keltirilgan, bu mavzuda kommunikativ til imkoniyatlarini rivojlantirishda korpusdan foydalandik.

Misol sifatida bitta mavzu keltirilgan, bu mavzuda kommunikativ til imkoniyatlarini rivojlantirishda korpusdan foydalandik.

Yuborish va qabul qilish

So‘z tuzilishi

aloqa qilmoq	aloqador, aloqadorlik,	ifodalamoq	ifodalash, ifodaladi
bashorat qilmoq	bashoratchi	maxfiy	maxfiylik, maxfiylashtirmoq
bog‘lamoq	bog‘langan, bog‘lanish	tarjima qilmoq	tarjima, tarjimon
dadil bo‘lmoq	dadillik	yetkazmoq	yetkazish, yetkazuvchi
gapirmoq	gapiruvchi, gapirdi, gapirgan	xabar qilmoq	xabardor, bexabar, boxabar

So‘z birikmalar

aloqani uzib qo‘ymoq
go‘shakni qo‘ymoq
internetga ulanmoq
internetdan foydalanishni to‘xtatmoq
kompyuterdan ma’lumotlarni qog‘ozga ko‘chirmoq
ma’lumot bilan to‘ldirmoq
qayta qo‘ng‘iroq qilmoq
yaqqol ma’lum bo‘lmoq

Mavzulashtirilgan lugʻat

ahamiyat (ot)	norasmiy (sifat)	telekoʻrsatuv va
aloqa (ot),	ommaviy axborot	radioeshittirish (ot)
aloqa oʻrnatmoq (feʼl)	vositalari (ot)	tomoshabin (ot)
aniq (sifat)	pichirlamoq (feʼl)	toʻxtamoq (feʼl)
aytmoq (feʼl)	qoʻngʻiroq qilmoq (feʼl)	toʻxtatmoq (feʼl)
bogʻlanmoq (feʼl)	rasmiy (sifat)	toʻgʻridan toʻgʻri (ravish)
eʼlon (ot)	ravshanlashmoq (feʼl)	veb sayt (ot)
hujjat (ot)	soʻz bermok (feʼl)	xabar (ot)
internet (ot)	talab qilmoq (feʼl)	xabar bermok (feʼl)
ishontirmok (feʼl)	tasvir (ot)	yozmoq (feʼl)
mobil telefon (ot)	telekanal (ot)	chop etmoq (feʼl)

Namuna sifatida keltirilgan mavzu bayonida, YUBORISH VA QABUL QILISH mavzusida oʻquvchi mavzulashtirilgan soʻzlar, yaʼni mavzuga doir boʻlgan soʻzlarning qaysi turkumda kelishini (OT, SIFAT, FEʼL, RAVISH) va ular ishtirokida gaplar tuzish strukturasini oʻrganadi. Soʻz birikmalari, soʻz tuzilishi, soʻzlardan namunalar qismida oʻquvchi soʻzlarni yod olib, matn tuzib, leksikonini boyitish imkoniyatiga ega boʻladi. Bunda kommunikativ til imkoniyatlarini rivojlantirish alohida ahamiyat kasb etadi

G Bosh harflarda berilgan soʻzlarning shaklini oʻzgartirib, gaplarni toʻldiring.

1. U juda kuchli ishonch bilan gapirdiki, soʻzlaridan ... yaqqol sezilib turar edi. **DADIL BOʻLMOQ**

2. Siyosatchilar chel el safariga chiqqanlarida koʻpincha ...ning yordamiga ehtiyoj sezadilar.

TARJIMA QILMOQ

3. Direktorning maktab taʼtili haqida ... gaplari majlisga yigʻilganlarning barchasiga birdek maʼqul boʻldi. **GAPIRMOQ**

4. Bugungi axborot zamonida istalgan kompyuterning internetga ... jarayoni oson kechmoqda.

BOGʻLAMOQ

5. Ijodkor oʻz mavzusiga ... boʻlgan barcha manbalarni izchillik bilan oʻrganib chiqadi. **ALOQA QILMOQ**

6. Insonlar hayotda baʼzan ... aytgan gaplarga butunlay ishonib qoladilar. **BASHORAT QILMOQ**

J Har ikkinchi gapni berilgan soʻzlardan foydalanib toʻldiring. Bunda birinchi gap bilan oʻxshash maʼno hosil boʻladi.

1. Nigora sening yangi veb sayting haqida nimadir dedimi? fikr bildirmok

Nigora sening yangi veb sayting ...?

2. Javohir menga oʻzining yangi mobil telefonini tasvirlab berdi. aytmoq

Javohir menga o'zining yangi mobil telefoni ...

3. Bu veb sayt tarixdagi ayrim aloqalarni tasvirlaydi. ma'lumot

Bu veb sayt tarixdagi ayrim aloqalar ... beradi.

4. Farhodga bizning kelgusi rejalarimiz haqida xat jo'natishni eslat! yozmoq

Farhodga bizning kelgusi rejalarimiz ... ni eslat!

Bu namunada biz o'quvchining muloqot kompetentligini rivojlantirish uchun mashqlardan foydalandik. Masalan o'quvchi bu platformada:

- ✓ **gap tuzish qoidalarini;**
- ✓ **zamonni farqlay oladi;**
- ✓ **mantiqiy fikrlashni o'rganadi;**
- ✓ **so'z birikmalarini bilib oladi;**
- ✓ **so'z tuzilishini;**
- ✓ **matn tuzish strukturasini qo'llab, amalda yoza oladi.**

Bundan ko'rinib turibdiki, global internet kompyuter tarmog'i ta'lim tizimini takomillashtirish uchun yangi istiqbollarni ochadi. O'zbek tili dunyoning rivojlangan tillari qatorida turadi. Shu bilan birga kompetensiyalardan foydalanish til imkoniyatlarini rivojlantirishda muhim ahamiyatga ega.

Ta'lim jarayonida o'zbek tilining kommunikativ kompetensiyasini shakllantirishga turli xil internet aloqa vositalari (elektron pochta, chat, forumlar, veb-konferensiyalar, internet loyihalari va boshqalar) ta'sir ko'rsatadi. Tillarni o'rgatishda o'qituvchi uchun qo'shimcha ma'lumot manbai sifatida internet saytlaridan keng foydalanish keng tus olayotganligini ta'kidlaydilar. Interaktiv, kommunikativ elektron ta'lim o'quvchilarda tanqidiy fikrlash va muammolarni hal qilish, muloqot, hamkorlik va ijodkorlik kabi ko'nikmalarni rivojlanishiga yordam beradi.

Foydalanilgan adabiyotlar:

1. Abdurahmonova N. A two-level morphological analysis of the Uzbek corpus /МАТЕРИАЛЫ IV международного научного конгресса иностранная филология. социальная и национальная вариативность языка и литературы / 2019/4/1 Симферополь, -P. 425-431
2. Abduraxmonova N. Til texnologiyasini rivojlantirishda o'zbek tili milliy korpusining innovatsion ahamiyati / "Yoshlar Innovatsion faolligini oshirishning dolzarb vazifalari" mavzusidagi Respublika ilmiy-amaliy konferensiya, Toshkent, 2019. – B. 159-162
3. Tawileh, W., Bukvova, H. & Schoop, E. (2013). Virtual hamkorlikda ta'lim: rivojlanayotgan mamlakatlar uchun Web 2.0 asosidagi elektron ta'limning imkoniyatlari va muammolari. N. Azab (Tahr.), Rivojlanayotgan mamlakatlarda Web 2.0 misollari: Amalga oshirish, qo'llash va foydalanish bo'yicha tadqiqotlar (380-410-betlar). Hershey, Pensilvanya: Axborot fanlari bo'yicha ma'lumotnoma.
4. <https://uzbekcorpus.uz/>
5. www.pedagog.uz
6. www.uzedu.uz

USING DIGITAL METADATA IN ELECTRONIC CORPORA

Arabboyeva Gulchehra,

Master's student

National university of Uzbekistan

Annotatsiya: Texnologiya rivojlangani sari ilmiy tadqiqotlar olib borishda izlanuvchilar yanada tezkorlik va aniqlikka intilmoqdalar. Vaqtni tejash va samarali ishlash uchun elektron bazalardagi turli omillardan foydalanishmoqda. Ushbu maqolada mana shunday omillardan biri bo'lgan raqamli metama'lumotlar haqida keng ma'lumot berilgan. Ularning vazifalari, foydali jihatlari va yaratilish jarayonlari atroflicha tahlil qilingan.

Kalit so'zlar: metama'lumot, Dubline Core, korpus.

Annotation: As technology advances, researchers are seeking more speed and accuracy in conducting scientific research. They are using various factors in electronic databases to save time and work efficiently. This article provides extensive information on one such factor, digital metadata. The tasks, useful aspects and creation processes of metadata are thoroughly analyzed.

Keywords: metadata, Dubline Core, corpus.

Аннотация: По мере развития технологий исследователи стремятся к большей скорости и точности проведения научных исследований. Они используют различные факторы в электронных базах данных, чтобы сэкономить время и работать эффективно. В этой статье представлена обширная информация об одном из таких факторов — цифровых метаданных. Тщательно анализируются задачи, полезные аспекты и процессы создания метаданных.

Ключевые слова: метаданные, Dubline Core, корпус.

Hozirgi kunda lingvistika va boshqa sohalarga aloqador tadqiqotlarni olib borish elektron manbalar va turli ma'lumotlarni o'zida jamlagan korpuslar orqali keng miqyosda amalga oshirilmoqda. Bunday elektron manbalardan foydalanish izalanuvchilarga bir qancha qulayliklarni taqdim etadi. Eng asosiy ana shunday qulayliklardan biri bu vaqtni tejashdir. Katta ma'lumotlarni o'zida jamlagan til korpuslaridan foydalanishda vaqtni tejovchi omillardan biri bu metama'lumotlardir.

Metama'lumotga ko'plab adabiyotlarda ma'lumotlar haqida ma'lumot deb ta'rif berilgan. Metama'lumot axborotni qayta ishlash sohasiga tegishli atama hisoblanadi. Yana ham aniqroq qilib tushintiradigan bo'lsak, meta'malumot bu biz o'rganayotgan manbaga aloqador bo'lgan qo'chimcha ma'lumotdir. Bu ma'lumot orqali biz nafaqat vaqtni tejashimiz, balki manbaning aslini yoki unga aloqador boshqa manbalarni topishimiz va ma'lumot aniqligiga ishonch hosil qilishimiz mumkin. Metama'lumotlar metadata deb ham yuritladi.

Metadata resurslarni topish, identifikatsiya qilish, resurslarni tashkil qilish va ma'lumotlarning o'zini almashish imkoniyatini, shuningdek, u taqdim etgan resurs yoki resurslarning almashinuvini osonlashtirish va qo'llab-quvvatlash uchun ishlatiladi. Metadata,

shuningdek, muhim kontekstli tafsilotlarni oladi va taqdim etadi, chunki hamma resurslar o'zini o'zi tavsiflay olmaydi.

Jahonda yaratilgan metama'lumotlar orasida Dublin Core Metadata alohida ajralib turadi. Dublin Core metadata elementlari to'plami hujjatlar, tasvirlar, videolar va boshqa turdagi resurslar kabi raqamli obyektlarni tavsiflash uchun standartdir. Dublin Core metama'lumotlari raqamli obyektlarning turli tomonlarini tavsiflash uchun ishlatilishi mumkin bo'lgan 15 ta elementdan iborat oddiy va ixcham to'plamni taqdim etadi.

Ushbu standartdan foydalanish raqamli resurslarning katta to'plamlarini boshqaradigan tashkilotlar uchun juda foydali. Dublin Core metadata elementlari to'plamiga rioya qilish orqali ushbu tashkilotlar o'z resurslarining to'g'ri tasvirlanganligini va osongina topilishini ta'minlashi mumkin. Bu foydalanuvchilar tomonidan yanada samarali qidirish va ko'rishni osonlashtirishi va raqamli kontentning vaqt o'tishi bilan to'g'ri boshqarilishi va saqlanishini ta'minlashga yordam beradi.

Dublin Core metama'lumotlaridan foydalanish orqali taqdim etiladigan ba'zi imkoniyatlarga qidiruv tizimini optimallashtirish, nogironlar uchun ko'proq kirish imkoniyati, resurslarni yanada samaraliroq topish va qidirish hamda aniqroq va izchil manba tavsiflari kiradi. Umuman olganda, Dublin Core metama'lumotlaridan foydalanish raqamli resurslarni yanada foydali va samaraliroq boshqarishga intiladigan har qanday tashkilot uchun qimmatli vositadir.

Metadata yaratishda bir qancha omillarni hisobga olinadi. Xususan, metadata yaratish uchun qanday ma'lumotlar kiritilishini aniqlash va keyin uni o'z ichiga olgan ma'lumotlar to'plamini yaratish kerak bo'ladi. Metadata yaratishning eng keng tarqalgan formati metadata sxemasi yoki Dublin Core, VRACore, ISO 19115 kabi standart yoki metadata yaratayotgan kontentga tegishli boshqa sxemadan foydalanishdir. Kiritilishi kerak bo'lgan ma'lumotlarni aniqlagandan so'ng, metadata elementlarini o'z ichiga olgan bir nechta yozuvlar yoki ma'lumotlar yozuvlarini yaratish kerak bo'ladi. Ma'lumotni aniqlash va topish oson bo'lishi uchun elementlar aniq belgilanishi kerak. Nihoyat, metadata yozuvlarini tegishli formatda, masalan, XML yoki JSON faylida saqlash maqsadga muvofiq bo'ladi.

Meta-ma'lumotlarni yaratishda quyidagilar muhim:

- aniqlik;
- to'liqlik;
- muvofiqlik;
- o'zaro ishlash:

Metama'lumot yaratayotganda to'plangan ma'lumotlar qaysi auditoriyaga tegishli ekanligi, resursni aniqlash va kontekstni to'g'ri shakllantirish uchun qanday ma'lumotlar kerakligini inobatga olish kerak bo'ladi. Bundan tashqari izlanuvchilar uni qanday topishlari, hozir va keyinchalik undan qanday foydalanishlari mumkinligi ham inobatga olinishi lozim.

Metadalar korpus turi va unga kiritilayotgan manbalarga qarab tadqiqotchilar tomonidan turlicha yaratiladi. Misol uchun gazeta korpusini yaratishda metadataga quyidagi elementlarni kiritish mumkin bo'ladi:

- nashr etilgan sana
- Sarlavha
- Manba
- Muallif

- Sahifa raqami
- Til yoki dialect
- Bo‘lim
- Mavzu yoki rukn
- Geografik joylashuv
- Kalit so‘zlar
- Nashr qilish chastotasi va shu kabilar.

Uzbekcorpus.uz bazasi tarkibida yaratilayotgan gazeta korpusi esa quyidagi jadvalda keltirilgan elementlarni o‘z ichiga olgan. Ushbu metadata Dublin Core va boshqa metadatalardan andoza olgan holda tuzib chiqilgan.

ID	Maqola nomi	Muallifi	Gazeta soni	Nashr etilgan sanasi	Rukni	Yozuv turi
1	Германия иқтисодий жамс “Дунё” АА, Берлин		121	2020 йил 2-июль	НИГОҲ	кирилл
2	Адолат тантанаси инсон та	Алишер Қодиров Олий Мажлис Қ	121	2020 йил 2-июль	ЖАРАЁН	кирилл
3	Ақш билан ҳамкорликнинг ғайрат Хонназаров, Ўза мухбири		121	2020 йил 2-июль	ЖАРАЁН	кирилл
4	Бўш турган объектлардан Ўзбекистон Республикаси Презид		121	2020 йил 2-июль	СИЁСАТ	кирилл
5	Миллий юксалиш мужасс Минҳожиддин Мирзо Республика		121	2020 йил 2-июль	2 ИЮЛЬ ЎЗБЕКИСТОН РЕСПУ	кирилл
6	Муסיқа инсон тафаккурини	Ойдин Абдуллаева, Олий Мажлис	121	2020 йил 2-июль	СИЁСАТ	кирилл
7	Ободлик бор жойда осойи Баҳор Хидирова, “Янги Ўзбекисто		121	2020 йил 2-июль	ЖАРАЁН	кирилл
8	Халқ ва давлат яқдиллигини Бекзод Наримонов Тошкент давл		121	2020 йил 2-июль	2 ИЮЛЬ ЎЗБЕКИСТОН РЕСПУ	кирилл
9	Аҳоли тармоқларидан фой	Ўзбекистон Республикаси Презид	121	2020 йил 2-июль	СИЁСАТ	кирилл
10	Эзгу интилишларимиз ва	Малика Қодирхонова Олий Мажл	121	2020 йил 2-июль	2 ИЮЛЬ ЎЗБЕКИСТОН РЕСПУ	кирилл
11	Аҳли донишлар аро ҳам р	Абдуллатиф Турдалиев Ғафур Ғу	121	2020 йил 2-июль	МАЪНАВИЯТ	кирилл
12	Жаҳон банкининг минтақа	Дилшод Раҳмонов Банк-молия ак	121	2020 йил 2-июль	НИГОҲ	кирилл
13	Жаҳон банкининг минтақа	Элмурод Ҳошимов Бизнес ва бош	121	2020 йил 2-июль	НИГОҲ	кирилл
14	Ўзбекистонлик тадқиқотчи	Назokat Абдуқундузова, Инновац	121	2020 йил 2-июль	ЖАРАЁН	кирилл

Gazeta korpusining metama’lumotlari ma’lum bir gazeta korpusida chop etilgan maqolalar, mualliflar va hikoyalar soni haqidagi statistik ma’lumotlarni o‘z ichiga oladi. Shuningdek, u maqolalarning o‘rtacha uzunligi, korpusda yoritilgan til(lar) va mavzular, nashr sanalari va gazeta tomonidan mo‘ljallangan geografik hudud va auditoriya kabi boshqa demografik ma’lumotlarni o‘z ichiga olishi mumkin. Bundan tashqari, metama’lumotlar tasvirlar va audio kabi multimediyadan foydalanish, shuningdek korpusga kiritilgan tashqi veb-saytlarga havolalar soni haqida ma’lumot berishi mumkin.

Ushbu metama’lumotlar tadqiqotchilarga gazeta korpusining mazmunini aniqlash, tavsiflash va tahlil qilish imkonini beradi. Bu ma’lumotlarni mavzular bo‘yicha tartibga solishga, tendensiyalar va aloqalar uchun ma’lumotlarni tahlil qilishga yordam beradi va qiyosiy tahlil uchun dalillarni taqdim etadi. Shuningdek, u gazeta korpusining kelib chiqishini va u yaratilgan kontekstni aniqlashga yordam beradi. Bundan tashqari, u muallif, nashr sanasi va korpusda ishlatiladigan til haqida ma’lumot berishi mumkin. Metadata tadqiqotchilarga gazeta korpusini chuqurroq tushunishga yordam beradi va ularga o‘z resurslaridan yaxshiroq foydalanishga imkon yaratadi.

Xulosa sifatida shuni aytish mumkinki, korporatsiyalar uchun metama’lumotlar ma’lumotlar manbai, ma’lumotlar turi, tili va boshqalar haqida qimmatli ma’lumot beradi. Bu ma’lumotlar tadqiqotchilarga tegishli korpuslarni topishga, ularning tadqiqot loyihasiga muvofiqligini baholashga va ulardan foydalanish kontekstini yaxshiroq tushunishga yordam beradi. Batafsil metama’lumotlarga ega bo‘lish tadqiqotchilarga kerakli ma’lumotlarga tez va

samarali kirishga yordam beradi va tajribalarni ko'paytirish va meta-tahlil o'tkazish uchun zarur poydevor yaratadi.

Foydalanilgan adabiyotlar:

1. SETIMES - A parallel corpus of the Balkan languages (2015). [Dataset (Text corpus)]. CLARIN:EL. <http://hdl.handle.net/11500/ATHENA-0000-0000-2591-2>
2. Parallel Global Voices. (2015). [Dataset (Text corpus)]. CLARIN:EL. <http://hdl.handle.net/11500/ATHENA-0000-0000-25DD-E>
3. Baca, M., & Getty Research Institute. (2008). Introduction to metadata. Los Angeles, CA: Getty Research Institute. Link: <http://d2aohiyo3d3idm.cloudfront.net/publications/virtuallibrary/0892368969.pdf>
4. Setting the stage / Anne J. Gilliland. Practical principles for metadata creation and maintenance Dublin Core Metadata Initiative. (2012). Dublin Core Metadata Element Set, Version 1.1. Link: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>
5. NISO Framework Advisory Group. (2007). A framework of guidance for building good digital collections. Bethesda, MD: NISO Press. Link: <http://www.niso.org/publications/rp/framework3.pdf> Metadata (pages 63-85)
6. Riley, J. (2017). Understanding metadata: What is metadata and what is it for? Baltimore, MD: National Information Standards Organization (NISO). Link: <https://www.niso.org/publications/understanding-metadata-2017>
7. 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 VI International Conference "Modern Problems of Applied Mathematics and Information Technology - Al-Khorezmiy 2018", pp. 37–38, Tashkent, Uzbekistan (2018)
8. Abduraxmonova, N. Z. «Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aforef.» (2018).
9. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
10. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
11. Abdurakhmonova N, Urdishiev 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.
12. 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;5(03):89-100.
13. 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
14. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.
 15. Abduraxmonova, N. Z., & Arabboyeva, G. S. (2022). Xronologik lingvistik korpus yaratishda gazeta matnlari o'rganish obyekti sifatida. Academic Research in Educational Sciences, 3(4), 663–667. <https://doi.org/10.24412/2181-1385-2022-4-663-667>

PARALLEL KORPUSNING O'RGANILISHIGA DOIR TAJRIBALAR TAHLILI

Anorboyeva Dilshoda Ibrohimjon qiziO'zbekiston Milliy universiteti Kompyuter lingvistikasi
va amaliy tilshunoslik mutaxassisligi 1-bosqich magistrantiE-mail: anorboyevadilshoda@gmail.com**Jo'raqulova Mushtariy**O'zbekiston Milliy universiteti Kompyuter lingvistikasi
va amaliy tilshunoslik mutaxassisligi 1-bosqich magistrantiE-mail: mushtariy00712@gmail.com

Annotatsiya: Hozirgi rivojlangan davlatlarda parallel korpuslarga doir bir qator tadqiqotlar qilingan bo'lib, kompyuter lingvistikasining rivojlanishida muhim manba hisoblanmoqda. Parallel korpus asosida matndagi turli elementlarning leksik, frazeologik, grammatik funksiyalarini tushunib yetish, ularga har tomonlama mos muqobillar topish keng qamrovli va murakkab jarayon hisoblanadi. Parallel korpus faqatgina mashina tarjimasi uchun lingvistik resurs bo'libgina qolmay, qiyosiy tarjima, avtomatik tarjima, chog'ishtirma tilshunoslik kabi sohalar uchun muhim hisoblanadi. Maqolada zamonaviy kompyuter texnologiyalar yordamida yaratilgan va yaratilini kelayotgan korpus va uning imkoniyatlarini takomillashtirishga oid fikrlar tahlilga tortilgan. O'zbek tili elektron korpusining ilmiy parallel korpuslarini yaratishda xorij tajribasi o'rganilib, o'zbek- ingliz tilining ilmiy bazasiga kiritishning amaliy jihatlariga e'tibor qaratilgan.

Kalit so'zlar: informatsion texnologiya, kompyuter lingvistikasi, korpus, matn korpusi, o'zbek tili korpusi, parallel matnlar korpusi.

Annotation: A lot of studies on parallel corpora have been conducted in modern developed countries, and they are considered a significant source in the development of computer linguistics. Understanding the lexical, phraseological, and grammatical functions of different elements in the text based on the parallel corpus, and finding suitable alternatives for them in all respects is a comprehensive and complex process. Parallel corpus is not only a linguistic resource for machine translation, but also important for fields such as comparative translation, automatic translation, hybrid linguistics. The article analyzes the corpus created and being created with the help of modern computer technologies and the ideas for improving its capabilities. The foreign experience in creating scientific parallel corpora of the electronic corpus of the Uzbek language is studied, and attention is paid to the practical aspects of the introduction of the Uzbek English language into the scientific base.

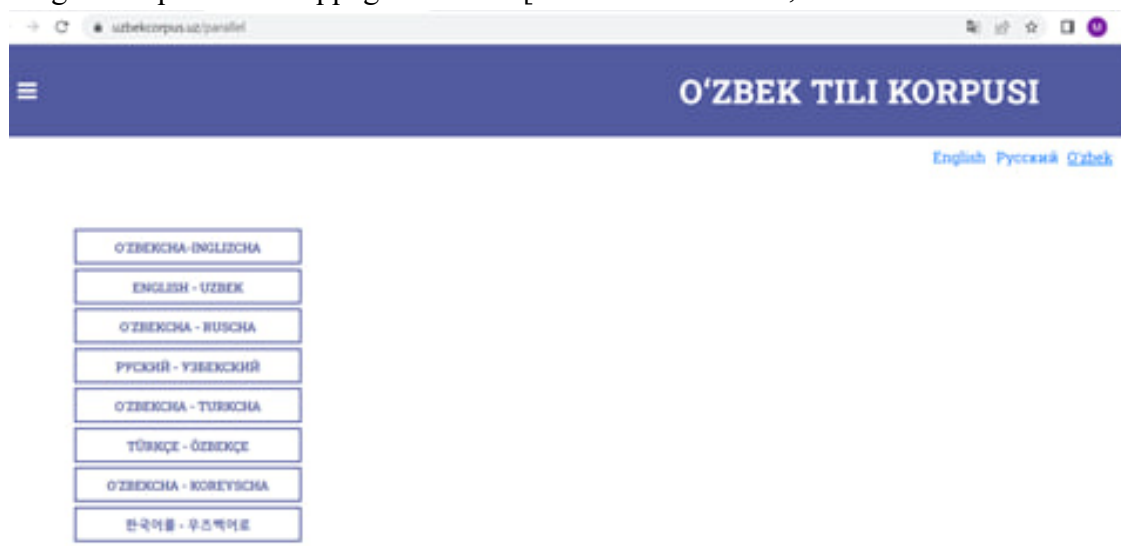
Keywords: information technology, computational linguistics, corpus, text corpus, national corpus, parallel corpus.

Аннотация: Ряд исследований параллельных корпусов был проведен в современных развитых странах, и они считаются важным источником в развитии компьютерной лингвистики. Понимание лексических, фразеологических и грамматических функций различных элементов в тексте на основе параллельного корпуса и нахождение для

них подходящих во всех отношениях альтернатив — комплексный и сложный процесс. Параллельный корпус является не только лингвистическим ресурсом для машинного перевода, но также важен для таких областей, как сравнительный перевод, автоматический перевод, гибридная лингвистика. В статье анализируется корпус, созданный и создаваемый с помощью современных компьютерных технологий, и идеи по улучшению его возможностей. Изучен зарубежный опыт создания научных параллельных корпусов электронного корпуса узбекского языка, а также уделено внимание практическим аспектам введения в научную базу узбекско-английского языка.

Ключевые слова: информационные технологии, компьютерная лингвистика, корпус, корпус текстов, корпус узбекского языка, корпус параллельных текстов.

Bugungi kun hayotimizda informatsion texnologiyalar va kompyuter alohida ahamiyatga ega. Barcha sohalar uchun kompyuter muhim vositaga aylanib ulgurdi. Buning natijasida yangi yo'nalishlar vujudga kela boshladi. Jumladan, tilshunoslik va kompyuter texnologiyalari kesishuvi natijasida kompyuter lingvistikasi fani vujudga keldi. Savol tug'ilishi mumkin: korpus, matn korpusi, parallel korpus o'zi nima? "Korpus" (korpus) lotincha so'z bo'lib, "tana" degan ma'noni bildiradi. "Korpus so'z, so'z birikmasi, grammatik shakllarni, so'z ma'nosini muayyan qidiruv tizimi orqali topishni anglatuvchi elektron ko'rinishdagi matnlar jamlanmasidir"[1]. Matnlar korpusi – ma'lum til birliklarini elektron holda saqlash, demakdir. Kompyuterda yaratilgan birinchi matnlar korpusi Braun korpusi sanaladi. "Milliy korpus bu milliy tilning o'ziga xos xususiyatlarini aks ettiradigan undan tilshunoslar, lug'atshunoslar, kompyuter lingvistlar, dasturchilar, tarjimonlar, jurnalistlar, olimlar, o'qituvchilar, noshirlar va boshqa kasb mutaxassislari tomonidan keng qo'llaniladigan korpusdir. Til korpusi- tilni o'rganish va undagi muammolarini hal qilishda foydalaniladi va ular turli xil elektron kutubxonalardan yig'ilgan bo'ladi"[2]. Tadqiqotchi N.Abdurahmonovanning korpusni "Grammatik jihatdan annotatsiyalashning avtomatik usullarini ishlab chiqish, mashina tarjimasida parallel korpus yaratishga doir qarashlari diqqatga sazovor" [Abduraxmonova N.,202





“Ko‘p yillar davomida olimlar va tadqiqotchilar tomonidan ilmiy hamkorlikda va katta loyihalar asosida keng ko‘lamli o‘zbek tilining elektron korpusi kashf etilganligi ma‘lum. Bu korpus o‘z ichiga quyidagi qismlarni oladi. Ular quyidagilar lingvistik lug‘atlar, web sahifalar, o‘zbek tilining morfologik bazasi, o‘quv adabiyotlarni hamda ilmiy, badiiy rasmiy uslubdagi adabiyotlardir” [2]. Parallel matnlar korpusi – badiiy asar, axborot vositalari nashrlari, qo‘llanmalar, turli xil hujjatlarning ikki yoki undan ortiq tillardagi elektron ko‘rinishi, deb atash mumkin. Bunday korpuslar yordamida biror so‘z , jumlaning yoki matnlarning turli tillardagi variantlarini bilish mumkin. Hozirgi kunda parallel korpus tillararo tadqiqotlar uchun juda foydali manba ekanligi barchamizga ma‘lum. So‘nggi yillarda tilshunoslikning turli xil sohalarida, xususan, tipologiya, tarjimashunoslik, leksikografiyada emperik asos bo‘ldi. Zamonaviy tilshunoslikning til o‘rganish obyekti va kompyuter yordamida tarjima qilishning dasturiy ishlanmasi bo‘lib xizmat qilmoqda. O‘rgangan tadqiqotimizda bengal tili bo‘yicha SUPara korpusigacha juda sanoqli korpuslar bo‘lgan. Biroq ushbu korpus ham yetarli darajada lingvistik bazaga ega, deb ayta olmaymiz. “EMILLE (Enabling Minority Language Engineering) korpusi Lankaster hamda Sheffield universitetlari olimlari tomonidan yaratilgan bo‘lib, texnik va lingvistik resurs jihatdan mukammalligi bilan avvalgi Bengal korpuslaridan ajralib turadi. Bu korpus 14 Janubiy Osiyo tillari uchun monolingual korpusni o‘z ichiga oladi va bu korpusning 5 ta ingliz parallel korpus turlari bor. EMILLIE Bengal parallel korpus bazasi 200 ta tag so‘zlardan iborat bo‘lgan matnlardan iborat. Bu korpusdan Buyuk Britaniya hukumatining konsullik yo‘riqnomalarida manba sifatida foydalanilgan. Yana bir parallel korpus sifatida OPUSni olishimiz mumkin. Ushbu korpus 60 tildagi parallel korpuslardan tashkil topgan bo‘lib, bu til korpus ochiq manbali hujjatlar va ularining tarjima manbalari bilan ishlaydi. KDE4 qo‘llanma matnlari Bengal parallel korpusi uchun o‘quv qo‘llanma matni vazifasini bajaradi” [Md. Abdullah Al Mumin etc, 2012.]

Parallel matnlar korpusi – badiiy asar, axborot vositalari nashrlari, qo‘llanmalar, turli xil hujjatlarning ikki yoki undan ortiq tillardagi elektron ko‘rinishi hisoblanadi. Bunday korpuslar yordamida biror so‘z , jumlaning yoki matnlarning turli tillardagi variantlarini bilib olamiz. Bundan kelib chiqadiki, parallel korpus orqali chet tilini o‘qitish mumkin. Parallel korpus manba

va maqsad tillari o'rtasida tuzilgan katta matnlarni o'z ichiga oladi. Ular statistik tahlil va gipoteza sinovini o'tkazish, hodisalarni tekshirish yoki ma'lum bir olamdagi lingvistik qoidalarni tasdiqlash uchun ishlatiladi. Bugungi kunda turli xil tillardan o'zbek tiliga asarlar tarjima qilinyapti, lug'atlar yaratilyapti. Turli xil til korpuslaridan, ko'pincha, tarjimonlar foydalanishadi. Bu ham parallel korpusning bir namunasi. Parallel korpusning qiymati uning hajmi va tillarning miqdori bilan belgilanadi. Schneider va Green ta'riflaganidek: "Parallel korpuslar tillar o'rtasidagi farqlar mavjud bo'lgan joylarni yaqinlashtirishga imkon beradi". Buning natijasida foydalanish qulayligi sababli parallel korpuslar qo'llanila boshlandi. Parallel korpus nazariy savollarga, shu jumladan, tarjima jarayoni va fikrni ikki yoki undan ortiq turli tillarda ifodalash yoki asl matn va tarjima matn xususiyatlarini solishtirish imkonini beradi.

Korpusning nazariy ahamiyati ikki tilning leksik va grammatik vositalarini qiyosiy tahlil qilish imkoniyatini yaratadi. Masalan, rus olimlari rus tili va xitoy tilini solishtirishadi. Xitoy va rus tillari birinchi navbatda morfologiyada sezilarli farqlarga ega. O'zining boy morfologiyasi bilan rus tilidan farqli o'laroq, xitoy tili butunlay morfologiyaga ega emas. Sintaksisda, ayniqsa, o'zaro murakkab qo'shma gaplarning yasashida sezilarli farqlar kuzatiladi. Korpusga asoslangan yondashuv Xitoyda yangi ilmiy va tadqiqot paradigmasini keltirib chiqardi, tarjimashunoslikka sezilarli hissa qo'shdi. Ikkita korpusning mavjudligi, ya'ni xitoycha tarjimalar korpusi va bir xil janr va mavzu sohasiga tegishli tarjima matnlari va asl matnlar o'rtasida tafovut mavjudligi yoki yo'qligini aniqlash imkonini beradi. Ko'pgina tadqiqotchilarning fikricha, parallel korpuslar ham manba, ham yangi tushunchalar beradi. Bir tilli korpuslarni o'rganish tadqiqotchilar, talabalarning tillar va ularning semantik va madaniy xususiyatlarini solishtirish imkonini beradi. Shuningdek, matnlarda parallel korpus tuzayotgan vaqtda, tili va qiyosiy matnlardan farqli o'laroq, madaniyatlararo munosabatni, manba matnlarning tarjima matnlariga aylantirishida uslub xususiyatini ham hisobga olish zarur. Parallel korpus uchun manba matnlari va ularni tarjimalarini moslashtirish, ya'ni jumla yoki so'zda ikkalasi o'rtasida bog'lanish hosil qilish muhim. Bular, asosan, tarjima uchun ishlatiladi. Ta'kidlash joizki, parallel korpuslar tarjimada tadqiqot va o'qitish uchun juda foydali. Shuningdek, tarjima jarayonini osonlashtirish maqsadida qidiruv tizimi ham moslashtirishgan. Parallel matn korpusini o'rganish lingvistik fanlarni o'rganishga yordam beradi. Korpus lingvistikasi metodlari, turlari, tahlillari amaliyoti o'zbek tilshunosligining turli sohalarida rivojlanmoqda.

Ko'pgina parallel korpuslarga foydalanish uchun qulay konkardanslar orqali kirish mumkin, bu esa tillararo hodisalarni o'rganishni sezilarli darajada osonlashtiradi. Bunday korpuslar ham til o'rgatish uchun boy materiallar manbayi hisoblanadi.

Foydalanilgan adabiyotlar:

1. <http://ruskorpora.ru>
2. <https://uzbekcorpus.uz/>
3. Abduraxmonova N. Z. Kompyuter lingvistikasi.()-Melbrun: Globe edit,2020
4. Md. Abdullah Al Mumin, Abu Awal Md. Shoeb, Md. Reza Selim and M. Zafar Iqbal. SUPara: A Balanced English-Bengali Parallel Corpus. Dept. of Computer Science and Engineering, Shahjalal University of Science and Technology, Sylhet, Bangladesh: UST Journal of Science and Technology, Vol. 16, No.2, 2012; P:46-51.Abduraxmonova N. Z. Kompyuter lingvistikasi.()-Melbrun: Globe edit,2020.

5. 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), Tashkent, Uzbekistan, 2021, pp. 1-4, doi: 10.1109/ICISCT52966.2021.9670043.
6. N. Z. Abdurakhmonova, A. S. Ismailov and D. Mengliev, «Developing NLP Tool for Linguistic Analysis of Turkic Languages,» 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation, 2022, pp. 1790-1793, doi: 10.1109/SIBIRCON56155.2022.10017049.
7. 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.
8. N. Abdurakhmonova, I. Alisher and R. Sayfulleyeva, «MorphUz: Morphological Analyzer for the Uzbek Language,» 2022 7th International Conference on Computer Science and Engineering (UBMK), Diyarbakir, Turkey, 2022, pp. 61-66, doi: 10.1109/UBMK55850.2022.9919579.

ILMIY USLUBDA SOHAVIY TERMINOLOGIYANI YARATISHDA VALENTLIK TUSHUNCHASI

Toshpo‘latova Dilshoda Karimjanovna

Namangan davlat universiteti

filologiya fakulteti 2-bosqich magistranti.

E-mail: dilshodatoshpolatova88@gmail.com

Annotatsiya: Ushbu maqolada tilshunoslik fanida valentlik tushunchasining qo‘llanilishi, valentlikning turlari va tushunchaga berilgan ta’riflar muhokamasi keltirildi. Shuningdek, huquqshunoslik faniga oid terminlar o‘zbek va ingliz tillari parallel korpusi orqali tahlil qilinadi. Terminlarning semantik strukturasi va sintaktik birikish imkoniyatlari o‘rganildi.

Kalit so‘zlar: valentlik, leksik valentlik, semantik valentlik, sintaktik valentlik, majburiy va fakultativ valentlik, termin, terminografiya.

Abstract: This article discusses the use of the concept of valence in the science of linguistics, the types of valence and the definitions given to the concept. Also, terms related to jurisprudence are analyzed through the parallel corpus of Uzbek and English languages. The semantic structure of the terms and the possibilities of syntactic combination were studied.

Keywords: valence, lexical valence, semantic valence, syntactic valence, obligatory and facultative valence, term, terminography.

Аннотация: В данной статье рассматривается использование понятия валентности в языкознании, виды валентности и определения, данные понятию. Также через параллельный корпус узбекского и английского языков анализируются термины, относящиеся к юриспруденции. Изучена семантическая структура терминов и возможности синтаксического сочетания.

Ключевые слова: валентность, лексическая валентность, семантическая валентность, синтаксическая валентность б обязательная и факультативная валентность, термин, терминография.

Mustaqil O‘zbekiston Respublikasining birinchi Prezidenti Islom Abdug‘aniyevich Karimov: “Biz dunyodagi taraqqiy topgan davlatlar safiga kirishni maqsad qilar ekanmiz, avvalambor, olimlarimiz kam o‘rganilgan, dolzarb muammolarni tadqiq qilishga kirishishi, ularning ilmi, obro‘si olamda mashhur bo‘lishi, jahon ilm-fanining yuksak mezonlariga mos bo‘lishi darkor. Biz kuchlarni birlashtirmasak, ilmiy izlanishlar samaradorligini oshirmasak erishgan yutuqlarimizni mustahkamlay olmaymiz” [Karimov I. A. 1994-y, 9 –iyul,-degan jumalari bilan zamonaviy tilshunoslik oldida turgan ulkan ishlar mavjudligini ta’kidlagan edilar.

Til tizimining asosiy birligi hisoblangan so‘zni har tomonlama o‘rganish hozirgi zamon tilshunosligining asosiy vazifalaridan biridir.

Valentlik tushunchasi aslida kimyoviy termin bo‘lib, uning zamonaviy tilshunoslikda qo‘llanilishi, so‘zlarning birikish imkoniyatlarini ko‘rsatib beradi.

Adabiyotlar tahlili va metodika:

“Valentlik” aslida kimyoviy atama bo‘lib, lotincha “valentia” so‘zidan olingan bo‘lib, “quvvat”, “kuch” degan ma‘noni anglatadi. Kimyo fanida bir element atomining boshqa element atomi bilan birika olish xususiyatini ifodalaydi. Masalan, kislorod (O) ikki valentli hisoblanadi chunki u ikkita vodorod (H) atomi bilan to‘yinadi va bu birikuv natijasida suv (H_2O) hosil bo‘ladi.

“Valentlik” tushunchasi tilshunoslikka o‘tgan asrning 40 –yillarida kirib keldi. Ilk marta S. D. Kasnelsonning 1948 –yilda yozilgan “О грамматической категории” asarida lingvistik termin sifatida qo‘llaniladi. Lingvistikada bu atama leksemaning boshqa leksema bilan birika olish qobiliyatini ko‘rsatish uchun ishlatiladi. Valentlik nazariyasi tildagi so‘zlarning semantik, sintaktik, funksional faolligini o‘rganishga asoslanib paydo bo‘ldi. So‘zlarning nutqda birikib kelishi mutlaq hodisa emas, balki nisbiydir. Chunki so‘zlar grammatik model taqazo etgan har qanday so‘zlarga emas, balki ulardan ba‘zilariga birikadi [Qo‘chqortoyev I., 1973:73]. Demak, so‘zlarning o‘zaro birikish imkoniyati chegaralangan bo‘lib, bu har bir so‘zning ma‘no imkoniyati bilan o‘lchanadi. Aniqrog‘i, birikuvchi so‘z sememalarining semalar tarkibida shu sememalarni biriktirish qobiliyatiga ega, har ikkala sememaga umumiy xos semantik komponentlar (klassesmalar) bo‘lib, semantik sintagma sememalari shu semalar orqali o‘zaro birikadi, mantiqiy munosabatga kirishadi [Апресян Ю. Д., 1974: 79]. Ko‘rinadiki, so‘zlar ba‘zi so‘zlar bilan semantik birikib, sintaktik konstruksiyani hosil qilsa, boshqa so‘zlar bilan faqat lisoniy fakt, nominativ birlik xarakteriga egaligi bilangina umumiylikni hosil qiladi, ya‘ni ular o‘zaro mantiqiy aloqaga kirishmaydi. Bunday munosabatga kirishganda ham, bu faqat sun‘iy, shaklan bo‘lib, obyektiv turlicha faktlar bilan tasdiqlanmaydi. Qiyoslang: semiz xolodilnik, to‘rtburchak velosiped, odamni o‘qimoq kabilar. [Rasulov R., 2020]. Lekin bu termin ustida xilma- xil qarashlar mavjud. So‘zlardagi bog‘lanishni ifodalash uchun G‘arbiy Yevropa tilshunosligiga L. Tenyer bu tushunchani olib kirdi. Olim valentlikni fe‘l o‘qiy olishi mumkin bo‘lgan aktantlar miqdori sifatida belgiladi. Ushbu olim qarashlari asosida fe‘llar quyidagicha guruhlanadi:

1. Avalent, ya‘ni shaxssiz fe‘llar: O‘qildi..
2. Bir valentli: U uxlaydi
3. Ikki valentli: U to‘pni tepadi.
4. Uch valentli: U unga kitob berdi.

Har qanday gap asosida fe‘l yotadi, deb hisoblovchi verbosentrik nazariya uchun yuqoridagi fikrlar asoslidir. Valentlik barcha so‘zlarga xos bo‘lsa-da, so‘z birikmalari qonuniyati, avvalambor, predikativli birliklarda ko‘rib chiqiladi. S.D. Kasnelsonning fikricha, “predikatning valentlik xususiyati gap qurimasida reallashadi, har bir predikat boshqa bo‘laklar uchun bo‘sh o‘rin ochadi” [Nebaraliyev J. 2012]. Lekin hozirgi tilshunoslikda predikat atamasi ostida faqat fe‘l emas, balki otlar ham tushunilgani va predikatlar materialiga ko‘ra fe‘l va ot predikatlarga bo‘lingani kabi, valentlik ham so‘z va tilning har qanday boshqa sath birliklarining boshqasi bilan birikish xususiyati deb tushunish keng tarqaldi. Shuningdek, valentlik tushunchasi dastlab leksik birliklarga nisbatan qo‘llanilib, “leksik valentlik” atamasi ostida qo‘llangan. Keyinchalik esa sintaktik birliklarga nisbatan ham qo‘llanilishi natijasida “sintaktik valentlik” tushunchasi hosil bo‘ladi. Demakki, tilshunoslikdagi boshqa tushunchalar qatori valentlik tushunchasi ham kengayib, o‘rganilish darajasi chuqurlashib bormoqda. Ilmiy rivojlanish natijasida shakliy va mazmuniy valentlik ham ajratila boshlandi.

Bogʻlanuvchi elementlardan qaysi biri birikuvchi va qaysi biri biriktirib oluvchi ekanligiga koʻra valentlik ikki turga ajratildi:

1. Faol valentlik
2. Nofaol(passiv) valentlik

Bundan tashqari, lisoniy birliklar valentliklarning yuzaga chiqishi zarurligi yoki zaruriy emasligi belgisiga koʻra quyidagicha tasniflanadi:

1. Majburiy (obligator) valentlik
2. Fakultativ valentlik

Valentlikning bunday turga boʻlinishi pozitsion sintaksisning taʼsiri natijasidir. Valentlikni bu jihatdan oʻrganish pozitsion sintaksisning rivoji uchun keng imkoniyat yaratadi. nPozitsion sintaksis talqiniga muvofiq, gap tarkibida markaziy oʻrin egallagan tayanch soʻz, yaʼni predikat qator pozitsiyalarga oʻrin ochadi.mBu pozitsiyalardan ayrimlari gapning mazmuniy-grammatik shakllanishi uchun toʻldirilishi zaruriy boʻlsa, boshqalari zaruriy boʻlmaydi, qoʻshimcha axborot beradi. Masalan, **Salima sumkasidan daftar oldi** gapida daftar soʻzi zaruriy, sumka soʻzi esa fakultativ valentlikni toʻldirib kelgan birliklar sanaladi. Leksik valentliklarni bevosita yuzaga chiqishi soʻz birikmalarini hosil qiladi. Shunday qilib, xuddi kimyoda birikma hosil qilish uchun muayyan miqdordagi atomlar talab qilingani kabi tilshunoslikda ham feʼl talab qilgan boʻlaklarni unga biriktirsakkina, grammatik jihatdan toʻgʻri gap hosil boʻlishi haqida nazariya yetakchilik qiladi.

Soʻnggi davrlardagi tilshunoslik tadqiqotlarida valentlik tushunchasi va nazariyasi dunyo tilshunoslari qatori, oʻzbek tilshunoslari eʼtiborini ham oʻziga tortmoqda.

Bu holatga lingvist Abramov “valentlik - grammatika va leksikologiya, sintaksis va semantika oʻrtasidagi farqlovchi holatni oʻzida namoyon qilishin bilan asoslanadi”,- deya taʼrif bergan boʻlsa, L. Tenyerning fikricha “... soʻzlarning boshqa soʻzlar bilan birikishi tushuniladi», V.V. Burlakova esa “ ... ikki valentligi dcganda soʻzlarning morfemalar asosida birikish usullari haqida fikr yuritadi”

Umuman olganda, valentlik nazariyasi tilning barcha sathlarida keng qoʻllanilmoqda: fonologik sathda [Abduazizov 2008, Juravlyov, 2004], morfologik sathda [Stepanova, Xelbig 1978], lcsik sathda [Rozensvich,1984; Kibordina 1979, 1975, 1976, 1988; Peshkovskiy, 1956; Burlakova, 1981 va boshqalar].

LB. Dolininaning fikricha: “Valentlik — lingvistik tahlilning bir sathda til birliklarining bir-biri bilan birikish imkoniyatlarini hisobga olgan holda, bu birikish imkoniyatida ularning ichki xususiyatlari boʻlib, buni birinchi navbatda uning sintaktik holati deb hisoblashga aytiladi” [Nebaraliyev J., 2012].

Valentlikni aniqlashda feʼlning lugʻatdagi maʼnosiga tayanish holatlari ham koʻp uchragan. 1968-yilda nemis tilshunoslari Xelbig va Shenkel tomonidan nemis tilidagi feʼllarning birinchi valentlik lugʻati yaratilishi nazariyaning yanada rivojlanishiga sabab boʻladi. Valentlik doirasida inglizcha komplementatsiyani tahlil qilishga bagʻishlangan birinchi tadqiqotlar Emons va Allerton ishlarida kuzatiladi. [T Herbst,1999]. Lekin, olimlar Stepanova va E.V. Rozovalarning monografik ishlarida hatto qardosh boʻlmagan tillar doirasida ham valentlikni oʻrganishda feʼlning gap tarkibida bildirgan maʼnosiga tayanish lozimligi keltiriladi. Chunki koʻpgina feʼllar koʻp maʼnolilik xususiyatiga ega boʻlib, barcha maʼnolar ham lugʻatda keltirilmasligi mumkin. Shuning uchun feʼlning kontekstda bildirgan maʼnosi asosida valentligi oʻrganiladi.

Tilshunos olim G. Helbrig csa oʻz tadqiqotlarida valentligini uch turga boʻladi:

1. Logik (mantiqiy) valentlik.
2. Sintaktik valentlik
3. Semantik valentlik

Logik valentlik-logik predmetlar bir necha bo'sh argumentlarga ega bo'lishi inobatga olinsa, semantik valentlik-fe'llar ushbu kontekstga ma'no jihatidan mos keladigan sintaktik birliklarni talab qiladi. Sintaktik valentlik deb bo'sh joylarni majburiy yoki fakultativ tarzda boshqa sintaktik birliklar bilan to'ldirilishig aytiladi.

Y.D. Apresyaning ta'kidlashicha " valentlik uch ko'rinishga ega: semantik, leksik, sintaktik. Semantik valentlik – so'zning bir aniq belgisi asosida boshqa lug'at birliklarning bog'lanishi bo'lsa, leksik valentlik aniq so'zlarning soniga qarab so'zlarning bog'lanishiga aytiladi. Sintaktik valentlikda ayrim lug'at birliklarining boshqa so'zlar bilan boshqaruvi yoki boshqarilish imkoniyatlari hisobga olinadi". [Nebaraliyev J., 2012]. Tomas Herbt's valentlikni quyidagicha izohlaydi: "Valentlik, birinchi navbatda, so'z bilan bog'langan sintaktik hodisa emas, balki sintaksisga tegishli bo'lgan leksikologik hodisa sifatida qaraladi " [T. Herbt's., 1999]. Olimning fikricha, valentlik nazariyasi jummalarni tahlil qilishga yondashadi, bunda ma'lum so'zlarning boshqa birliklar paydo bo'lishi uchun zarurligi, o'ynaydigan rolga e'tabor qaratiladi.

Demak, valentlik nazariyasining tilshunoslar o'rtasida turlicha talqin qilinishi bu nazariyaning bahs talab masalalardan biri ekanligini namoyon etadi. Valentlikni til sathlarida farqlash masalasini Y.D. Aprcsyan o'rtaga tashlab, u semantik valentlikni sintaktik valentlikdan farqlashni ta'kidlaydi.

O'zbek tilshunosligida valentlikning quyidagi uch turi farqlanadi:

- 1) sintaktik valentlik; 2) semantik valentlik; 3) leksik valentlik.

Ushbu qarashda hokim va tobe bo'lak o'rtasidagi grammatik bog'lanish (sintaktik valentlik), so'zlar orasidagi mazmunli bog'lanish (semantik valentlik) va bir sinonimik qatoriga mansub so'zlardan birini uslubiy jihatdan bog'lanish (leksik valentlik) o'rni haqida nazariya ilgari surilgan. Ammo semantik valentlik va leksik valentlik orasida shunday bir umumiylik mavjudki, ularning alohida-alohida olinishiga to'sqinlik qiladi. Chunki birida so'zlar orasidagi mazmungan moslik asos qilib olinsa, ikkinchisida mazmun va uslub asos qilib olinadi. Uslubning zamirida ham aslida mazmun rol o'ynaydi: ijobiy va salbiylik nuqtayi-nazaridan, albatta, ba'zi bir sinonimlarda esa, uslub, umuman, farq qilmaydi: kulgu-xanda

Lug'aviy valentlik biriktiruvchi leksemaning ma'noviy jihatdan o'ziga mos birikuvchilarni tanlashi, deb izohlanadi. Masalan, *sayra* leksemasi *kuchuk* leksemasini o'ziga torta olmaydi. Chunki ularda ma'noviy muvofiqlik yo'q. Leksemalar bir- biriga mos bo'lishi uchun ularning ma'no strukturalarida o'xshash ma'no bo'lakchalari-sememalar bo'lmog'i lozim. Yana *non* va *yemoq* leksemalari ma'nolari shu ikki so'zning birikishini taqozo etadi: *non yemoq*, ammo *non* va *ichmoq* leksemalari ma'noviy jihatdan birika olmaydi, chunki so'zlarning semik tarkibi bunga yo'l qo'ymaydi.

Yuqoridagi hollardan ma'lumki, valentlikning bir turi shakl jihatidan bog'lanish imkoniyatini yuzaga chiqarsa, yana biri mazmun jihatidan bog'lanishni vujudga keltiradi.

Valentlik tahlil metodi so'z muammosi bilan shug'ullanadi, leksemaning nutqiy faoliyatdagi semantik, sintaktik va funksional faolligini o'rganadi, tekshiradi, uning ichki mazmun jihati bilan bog'liq imkoniyatlarini ochib beradi. Aniqrog'i, valentlik tahlil metodi orqali muayyan

soʻzning oʻz maʼnosi asosida boshqa soʻzlar bilan sintaktik (semantik- sintaktik) munosabatga kirishuvi aniqlanadi, yaʼni soʻzning turli maʼnoga ega qator soʻzlarni biriktirish xususiyati yotadi.

Tilshunoslikda ham, terminologiyada ham valentlik masalasi oʻrganiladi. Tilshunoslikda bu masala keng va atroflicha oʻrganilgan, lekin terminologiyada D.S. Lotte taʼbiri bilan aytganda, faqat muammo sifatida qoʻyilgan, juda zaif oʻrganilgan mavzu hisoblanadi [Ефимов Д. К., 2005]. Tadqiqotimizda terminlarning valentligi – boshqa soʻzlar bilan boʻlganish qobiliyati potensial deb olinadi. Tadqiqotimizning materiali huquqshunoslik yoʻnalishi maxsus matnlarining asl va tarjima nusxalari boʻlib, biz ushbu parallel matnlardagi atamalar xususida toʻxtalamiz. Oʻzbek va ingliz tillaridagi ilmiy atamalarni toʻplash va ikki tilli terminologik lugʻat bazasini shakllantirishga harakat qildik. Muayyan til terminologiyasining rivojlanishi jamiyat hayoti va unda sodir boʻlayotgan oʻzgarishlarga bogʻliq boʻlishi barchaga ayon va bu huquqshunoslik sohasiga ham taalluqli. Qonun va huquq tili, huquqiy terminologiya professional bilimni talab qiladi. Bu borada H. A. Kerimov shunday fikr bildiradi: “Notoʻgʻri yoki oʻrinsiz qoʻllangan soʻz yoki ibora fikr va uning matn ifodasi oʻrtasidagi tafovut baʼzan ogʻir oqibatlarga olib kelishi mumkin. Ayniqsa, qonun ijodkorligi sohasida” [Керимов Д. А., 1998]. Parallel matnlar korpusida, tarjima nazariyasida bu fikrlar katta ahamiyat kasb etadi. Atamalar nafaqat soʻz, balki, birikma shaklida boʻlishi ham mumkin. Soha doirasida terminlarni tuzilishiga koʻra quyidagi guruhlariga boʻlib olindi:

1. Sodda terminlar. *Sud- court, ayblov-accusation, bitim- deal, vakolat- authority, vasiylik-guardianship, garov-pledge, guvoh- witness, jabrlanuvchi- the victim, jazo- punishment, kafillik- surety, majburiyat- commitment, sanksiya - sanction, konstitutsiya -the constitution, burch- duty, huquq- right, advokat-lawyer, amnistiya- amnesty, ariza - application, autentlik-authenticity, soliq - tax, prokuror- prosecutor, kodeks- law-book, prokuratura-prosecutor's office.*

Yuqorida keltirilgan misollar orqali koʻrish mumkinki, oʻzbek tilida sodda shaklga ega boʻlgan termin ingliz tilidagi varianti qoʻshma shaklda ham boʻlishi mumkin: vasiylik - **guardianship: guardian – vasiy, ship- kema** (ushbu termin soʻzma –soʻz tarjimada ekvivalent boʻla olmaydi); jazo- **punishment: punish- jazolash, ment – yakka holda hech qanday maʼno ifodalamaydi; kodeks- law-book: law- qonun, huquq, book - kitob; prokuratura- prosecutor's office: prosecutor's –procurorning, office –idora.**

2. Qoʻshma terminlar: *ayirboshlash-exchange, shartnoma-contract, bojxona – customs, huquqbuzarlik-offence, afv etish-pardon, axloqbuzar-immoral;* Qoʻshma terminlar parallelligida ham nomuvofiqlik koʻzga tashlanadi. Oʻzbek tilida qoʻshma shaklda boʻlgan terminlar ingliz tilida sodda shaklda boʻlishi holati kuzatiladi.
3. Birikma terminlar: *boshqaruv shakli-management form, davlat nazorati – state control, daʼvo muddati-claim period, mehnat daftarchasi-work book, mualliflik huquqi- copyright, sinov muddati-trial period, fuqarolik jamiyati-civil society, gumon qilinuvchi- suspect, jinoiy javobgarlik- criminal liability, jismoniy shaxs-physical person, yuridik shaxs-legal entity, intellektual mulk-intellectual property, mansabdor shaxs- official, moddiy javobgarlik-material liability, yuridik akt-legal act, huquqiy layoqat-legal capacity.*

Birikma shaklida keltirilgan huquqiy terminlar parallel korpusda bir- biriga mos va mos boʻlmagan shakllarda uchraydi.

a) mos kelgan shakl: boshqaruv shakli-**management form**, davlat nazorati-**state control**, jismoniy shaxs-**physical person** kabi;

b) Mos bo‘lmagan shakl:

- Shakl jihatdan nomuvofiqlik: mansabdor shaxs- **official**, gumon qilinuvchi –**suspect**;
- Ma’no jihatdan nomuvofiqlik: moddiy javobgarlik –**material liability**, yuridik shaxs – **legal entity** –legal-huquqiy, entity tashkilot(ma’no jihatdan to‘g‘ri kelsa ham so‘zma so‘z tarjimada ekvivalent emas), mualliflik huquqi- **copyright**- copy- nusxa ko‘chiruvchi, right- to‘g‘ri (tarjima variantida so‘z ma’nosi butunlay o‘zgarib ketadi), mehnat daftarchasi –**work book** – work –mehnat, ish, book –kitob (“mehnat kitobi” shaklida tarjima)

4. Murakkab terminlar: **axloq tuzatish kaloniyalari-correctional colonies, bir taraflama bitim-one sided deal, bir umrga ozodlikdan mahrum etish- life imprisonment , normativ huquqiy akt-normative legal act, fuqaroning muomala layoqati-citizen`s legal capacity, ichki ishlar organlari –internal affairs;**

Yuqoridagi misollardan ko‘rishimiz mumkinki, huquqiy terminlarni parallellik nuqtayi nazaridan tahlil qilish murakkab ishlardan hisoblanadi. Terminlar parallel korpusda shakl va mazmun jihatdan bir –biriga ekvivalent bo‘lishi ham bo‘lmasligi ham mumkinligi kuzatildi. Shuningdek, bir nechta ingliz atamaları bitta o‘zbekcha atamaga mos bo‘lishi ham mumkin. Ko‘p so‘zli atamalar (multiword terms) o‘zgaruvchanlikka moyil bo‘lib, ularni boshqa tillarda ifodalash oson emas.

Tilshunoslik sohasida terminlarning o‘zgarishi va yangi terminlarning paydo bo‘lishi fan va texnika taraqqiyoti bilan uzviy bog‘liqdir. Kelib chiqishi jihatidan terminlar bir tilning ichki imkoniyatlari asosida paydo bo‘lishi yoki boshqa tillardan o‘zlashishi mumkin. Masalan: **qonuniy, huquqiy, ichki ishlar organlari fuqaroning muomala layoqati** kabi terminlar tilning ichki imkoniyatlari asosida paydo bo‘lgan bo‘lsa, **sanksiya, konstitutsiya, amnistiya** kabi birliklar tilimizga tushuncha bilan birgalikda kirib kelgan.

Quyidagi jadvalda “jinoyat” so‘ziga birika oladigan yuridik terminlarni keltirishga harakat qildik:

Jinoyat	Kodeksi
	Qidiruvi
	Protsessi
	protsessual immunitet
	protsessual kodeksi
	protsessual huquqi
	Tarkibi
	ijroiya kedeksi

Terminlarni o‘rganish jarayonida shartli ravishda standart va nostandart guruhga bo‘ldik. Standart terminlarning doimo nutq tarkibida ham, nutqdan tashqari holatda ham “yadroviy” semasi bitta bo‘ladi.

Masalan: **yuridik shaxs, legal entity, konstitutsiya – constitution**

Nostandart terminlar esa faqatgina o‘ziga tegishli matn tarkibidagina ilmiy terminligini ko‘rsatadi. Semantik va shakl xossalariga ega bo‘lmay, ma’lum bir terminologik tizim doirasida bo‘lgan holda bu xususiyatlarni “vaqtinchalik” egallaydi [Ефимов Д. К.,2005].

Masalan: bolaning mehnat qilish huquqi, boshqarish kabilar.

Terminlar soʻzma –soʻz tarjima qilinganda ekvivalent boʻlishi va boʻlmasligi mumkin. Masalan: *ichki ishlar organlari- internal affairs* (“*organ*” soʻzi tarjimada keltirilmaydi.)

Terminologik frazeologiya, odatda, taqqoslangan tillarda mos kelmaydi. Toʻplangan huquqshunoslik ilmiy matnlar terminologiyasida birikmali terminlar asosiy qismni tashkil etadi.

XULOSA. Yuqoridagi fikrlardan maʼlum boʻldiki, soʻz valentligi ikki tomonlama aloqa faolligi orqaligina amalga oshadi. Faqat hokim soʻz bilan emas, tobe soʻz bilan ham valentlik munosabati belgilanadi. Har qanday hokim soʻzning semantik qurilishida tobe soʻz yashirin boʻlsa ham mavjud boʻladi. Nutq jarayonida bu soʻz, yaʼni ot harakatga kelib, oʻziga loyiq grammatik shakllarni olgan sintaktik qurilmani yuzaga chiqaradi.

Foydalanilgan adabiyotlar:

1. Karimov I. A. Oʻzbekiston fanlar akademiyasi umumiy yigilishida soʻzlagan nutqidan. “Oʻzbekiston ovozi”, 1994-y, 9 -iyun
2. Qoʻchqortoyev I. Soʻz valentligi // Oʻzbek tili va adabiyoti. j.1973,35-bet
3. Апресян Ю. Д. Лексическая семантика Синонимические средства языка. М;1974. 79- 6
4. Rasulov R. Valentlik nazariyasi // Academic Research in Educational Sciences, Volume 3. 2022
5. Nebaraliyev J. Ingliz tili gap qurilmasida feʼl valentligi . BKM, Toshkent-2012
6. Nebaraliyev J. Ingliz tili gap qurilmasida feʼl valentligi . BKM, Toshkent-2012
7. T Herbst - Erfurt Electronic Studies in English, 1999
8. Nebaraliyev J. Ingliz tili gap qurilmasida feʼl valentligi . BKM, Toshkent-2012
9. T Herbst - Erfurt Electronic Studies in English, 1999
10. Ефимов Д. К. Терминосфера сми в русском и английском языках.(автореферат) Екатеринбург-2005
11. Керимов Д. А. Законотетельная техника:научно-метод. и учеб. Пособие.М: Норма-инфра –М 1998.
12. Ефимов Д. К. Терминосфера сми в русском и английском языках.(автореферат) Екатеринбург-2005

**O‘ZBEK KORPUS LINGVISTIKASIDA XI-XIX ASR ADABIY MANBALARI
LEKSIKOGRAFIYASINING O‘RNI****Qo‘ziyeva Sitora Ibodulla qizi**

Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti

O‘zbek filologiyasi 2-bosqich talabasi

E-mail: nkosmos1199@gmail.com

Ilmiy rahbar:

Nilufar Zaynobiddin qizi Abdurahmonova

Filologiya fanlari bo‘yicha falsafa doktori(f.f.d)

E-mail: abdurahmonova.1987@mail.ru

Annotatsiya: Bugungi kunda o‘zbek tili korpusi masalasi o‘zbek kompyuter lingvistikasida dolzarb masalalaridan biri bo‘lib qolmoqda. Ushbu tezisda korpus lingvistikasi boyishiga xizmat qiladigan XI-XIX asr adabiy manbalarining leksikografiyasi ustida olib borilayotgan kichik tadqiqot ishi haqida qisqacha yoritiladi.

Kalit so‘zlar: Kompyuter leksikografiyasi, korpus, korpus lingvistikasi, o‘zbek tili, morfologik ma’lumotlar bazasi, morfologik tahlil.

Abstract: Today, the issue of the Uzbek language corpus remains one of the urgent issues in Uzbek computer linguistics. In this thesis, a small research work on the lexicography of literary sources of the XI-XIX centuries, which serves to enrich corpus linguistics, is briefly reflected.

Keywords: Computer lexicography, corpus, corpus linguistics, Uzbek language, morphological database, morphological analysis.

Аннотация: На сегодняшний день проблема корпуса узбекского языка остается одной из актуальных проблем узбекской компьютерной лингвистики. В данной диссертации будет кратко освещена небольшая исследовательская работа по лексикографии литературных источников XI-XIX вв., служащая для обогащения корпусной лингвистики.

Ключевая слова: Компьютерная лексикография, корпус, корпусная лингвистика, узбекский язык, морфологическая база данных, морфологический анализ.

Kompyuter texnologiyalari va metodlari yordamida tabiiy tillarni qayta ishlashning lingvistik asoslarini yaratish hamda o‘zbek tili kompyuter lingvistikasiga tegishli nazariy va amaliy masalalarni o‘rganish muhimdir. Tabiiy tilning lisoniy modellari va nutqiy imkoniyatlarini kompyuter tiliga o‘tkazish bugungi axborot asrida kompyuter lingvistikasining asosiy masalalaridan biri bo‘lib qolmoqda. Va bu borada olib borilayotgan ishlarning ijobiy natijalari o‘laroq ko‘plab ilmiy o‘zgarishlar va yutuqlarga erishildi. Buning natijasida mashina tarjimas, korpus lingvistikasi, kompyuter leksikografiyasi, matnlarni tasniflash va shu kabi bir qator yo‘nalishlar vujudga keldi [Abdurahmonova N. 2021:6].

Jumladan, Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti kompyuter lingvistikasi va amaliy tilshunoslik kafedrasini mudiri, filologiya fanlari bo'yicha falsafa doktori Nilufar Abdurahmonovaning o'zbek tili korpusini yaratishdagi ishlari diqqatga sazovordir. Darhaqiqat, bu korpus O'zbekistonda birinchi bo'ldi. O'zbek tilining ushbu elektron korpusi bir necha yillik olib borilgan tadqiqotlar va 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'minot hisoblanadi. Hamda uning lingvistik ma'lumotlar bazasi boyishi uchun materiallar to'plash va uni kiritish kerak bo'ladi.

Shu tufayli biz kichik tadqiqot ishi sifatida o'zbek milliy nasridan leksikografik materiallar to'ladik. Ushbu ishdan ko'zlangan maqsad o'zbek korpusiga XI-XIX asrlar leksikografiyasini kiritish bo'lib, shu davr adabiy manbalari, jumladan, XI asrga doir manbalardan Mahmud Koshg'ariy, Yusuf Xos Hojib, Ahmad Yugnakiy, Ahmad Yassaviy, XIII-XIV asrga oid Nosiriddin Rabg'uziy, Pahlavon Mahmud, Sayfi Saroyi, Xorazmiy, XIV-XV asrlarga oid Sakkokiy, Atoiy, Lutfiy, Navoiy kabi ijodkorlarning ba'zi asarlaridan o'sha davrga oid tarixiy va bugungi kunda arxaizm elementiga kiradigan so'zlarni yig'dik.

Jarayonda muayyan davrga oid muayyan asardagi leksemalar fonetik, morfologik (asosan, morfologik, ya'ni so'z turkumlari bo'yicha), morfemik hamda lug'aviy tomondan o'rganildi. XV asrga oid yozma manbalardan Alisher Navoiyning "Hayrat-ul abror" dostoni misolida ko'rib chiqaylik. Bu asardan fe'llar yig'ib chiqilganda, ularning aksariyati bugungi sinxron aspektidagi tilimiz fe'llaridan deyarli farq qilmasligining guvohi bo'ldik. Misol uchun, asarda bugungi adabiy tilimizdagi *bo'lmoq* fe'lining *bo'lub*, *o'lg'ay*, *bo'lg'ay*, *bo'lubon*, *o'lsa*, *o'lub*, *bo'lg'usi*, *bo'lg'uchi*, *bo'lsa*, *o'lmoqki*, *bo'lg'ali*, *bo'lg'ucha*, *o'lubmen*, *bo'lsavu* kabi shakllari uchraydi. Morfologik tahlil doirasida bu so'zlarda asosdan so'ng qo'shilgan grammatik vositalar bugungi kundan olib qaralganda diaxron lingvistika tarkibida o'rganilishini e'tiborga oladigan bo'lsak, ushbu til birliklari ilm ahli uchun o'zbek korpusi lingvistikasida tarixiy asarlar grammatikasini o'rganishga ham zamin yaratadi. *Xatm o'lmoq*, *jazm aylamoq*, *zohir o'lmoq* kabi bir qator fe'llarni lug'aviy tomondan tahlilga tortadigan bo'lsak, ular bugungi tugamoq; bel bog'lamoq, qaror qilmoq, muqarrar qilmoq; paydo bo'lmoq, ko'rinmoq fe'llariga to'g'ri keladi. Undan tashqari asarda uchraydigan *qilmoq* fe'lining *qilg'ach*, *qilg'uchi*, *qilg'onidin*, *qilg'on*, *qilg'oni*, *qilg'usi* kabi ko'rinishlarini; kelmoq fe'lining *kelibon*, *kelmish*, *keldi*, *kelur*; *topmoq* fe'lining *topmag'ay*, *topilmas*, *topqay*, *topibon*, *topib*, *topmag'on*, *topti*; *aylamoq* fe'lining *aylabdur*, *aylasa*, *ayladim*, *aylagay*, *aylagil*, *aylagusi*, *aylabon*, *aylagaysen*, *aylamak* kabi ko'rinishlarini va h.k.o fe'llarning bugun deyarli uchramaydigan turli morfologik vositalar yordamida shakllanganini ko'rishimiz mumkin.

Xuddi shunday tahlillar bilan yana ot, sifat, son, olmosh, ravish hamda yordamchi so'zlar ham o'rganilmoqda. Kelajakda ushbu materiallar, lingvistik korpusdagi badiiy, ilmiy adabiyotlarning yirik leksikografik bazasi mobil ilova shaklida ommaga taqdim qilish maqsad qilingan! Bu, albatta, ixcham va qulaylikni ta'minlaydigan kutubxona vazifasini o'taydi, degan fikrlar ham yo'q emas.

Iste'moldan chiqish xavfi ostiga kelib qolgan ayrim leksemalar uchun ularning elektron korpusini yaratish, shu leksema tiliga taalluqli bo'lgan ilmiy va badiiy adabiyotlar

leksikografiyasini tiklash va kengaytirishga, uni yillar davomida asrashga, ular ustida bir qancha ilmiy ishlar qilishga imkon beradi.

Umuman olganda, lingvistik korpuslar yaratish ta'limga, ayniqsa yosh tadqiqotchilar ishlari samaradorligiga yordam beradi. Sababi til ijtimoiy hodisa sifatida doimiy o'zgarishda, qaysidir so'zlar neologizm sifatida kirib kelsa, ba'zilar esa tarixiy so'zlarga aylanadi. Bu jarayonni esa elektron korpuslar orqali bevosita kuzatib borish mumkin. Ko'rinib turibdiki, korpus nafaqat soha kishilari uchun asosiy mohiyat kasb etadi, balki bu tilni rivojlantirishda umummilliy masala ham hisoblanadi.

Foydalanilgan adabiyotlar:

1. Abdurahmonova N. Kompyuter lingvistikasi (darslik), Toshkent.: Akademnashr, 2021. – B11.
2. Abdurahmonova N. O'zbek tili elektron korpusining kompyuter modellari (monografiya),
3. Abdurahmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aforef." (2018).
4. Abdurahmonova, N. Z. O'zbek tili elektron korpusida (<http://uzbekcorpus.uz/>) og'zaki matnlar korpusini yaratishning nazariy va amaliy masalalari (maqola), 2022, – B7.
5. <http://uzbekcorpus.uz/>.

UZBEK ELECTRONIC CORPUS OF NEWSPAPER TEXTS

Mukhammadiyeva Kamola¹,
Mukhammadjonova Mavluda¹,
Ravshanova Dinora¹,
Sharipova Nafosat¹,

¹National University of Uzbekistan
Teaching philology and languages:
uzbek language 2nd year students

Annotation: These days, creating a corpus of newspaper texts is an urgent issue. The article describes the research on the creation of newspaper corpora, the linguistic research of the texts of the newspaper «Ma'rifat» published in the period of 2018-2023 included in the corpus of the Uzbek language. In addition, the texts in the corpus are linguistically annotated, their metadata is cited. Implementation of these newspaper texts in the electronic corpus of the Uzbek language reflects the fact that it will serve as a learning object for linguistic research in the future. Various studies on the creation of a corpus of newspaper texts in the world and their importance in linguistics are analyzed in detail. It is discussed about its special features and functions. It is also discussed about the researches made in Uzbekistan on a creating a corpus of newspaper texts. In addition, there is information about the creation of a corpus of newspaper texts and its advantages.

Keywords: Uzbek language electronic corpus, newspaper corpus, corpus linguistics, newspaper texts, corpora, multilingual, monolingual, infrastructure

Аннотация: На сегодняшний день создание корпуса газетных текстов является актуальным вопросом. В статье описаны исследования по созданию газетных корпусов, лингвистическое исследование текстов газеты «Ма'рифат», изданных в период 2018-2023 гг., включенных в корпус узбекского языка. Кроме того, тексты в корпусе лингвистически аннотированы, приводятся их метаданные. Внедрение этих газетных текстов в электронный корпус узбекского языка отражает тот факт, что в будущем он послужит учебным объектом для лингвистических исследований. Подробно анализируются различные исследования по созданию корпуса газетных текстов в мире и их значение в языкознании. Обсуждается его особенности и функции. Также говорится о проведенных в Узбекистане исследованиях по созданию корпуса газетных текстов. Кроме того, есть информация о создании корпуса газетных текстов и его преимуществах.

Ключевые слова: электронный корпус узбекского языка, газетный корпус, корпусная лингвистика, газетные тексты, корпусы, многоязычие, одноязычие, инфраструктура.

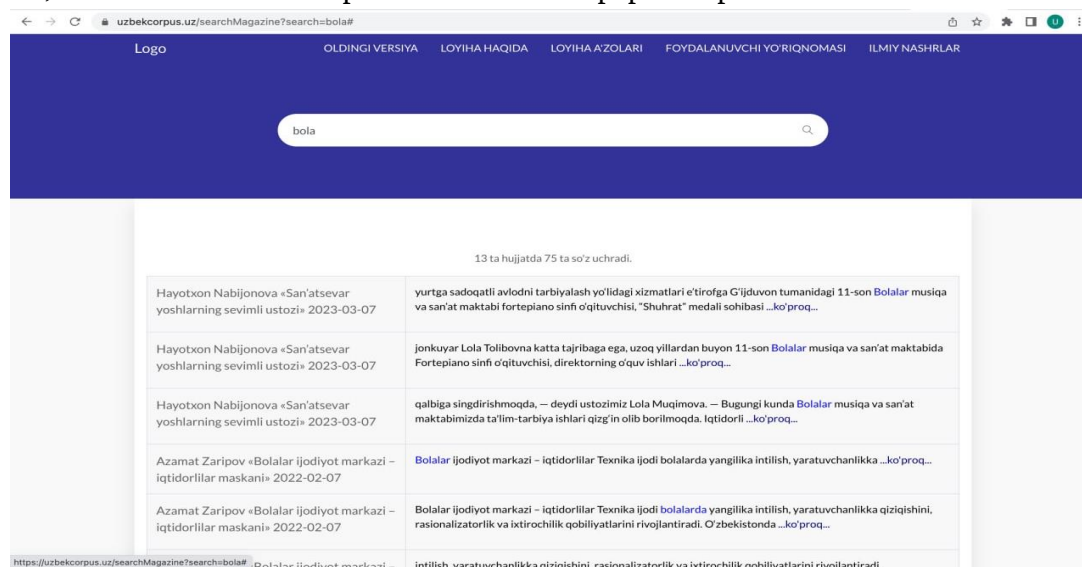
Introduction.

In many countries of the world, many researches and studies have been carried out on corpus linguistics, but in Uzbekistan, the initial work on corpus linguistics has been conducted and is now developing. In linguistics, a corpus or text corpus is a language resource consisting of a large and structured collection of texts. Corpus linguistics is a corpus created for linguistic. Some

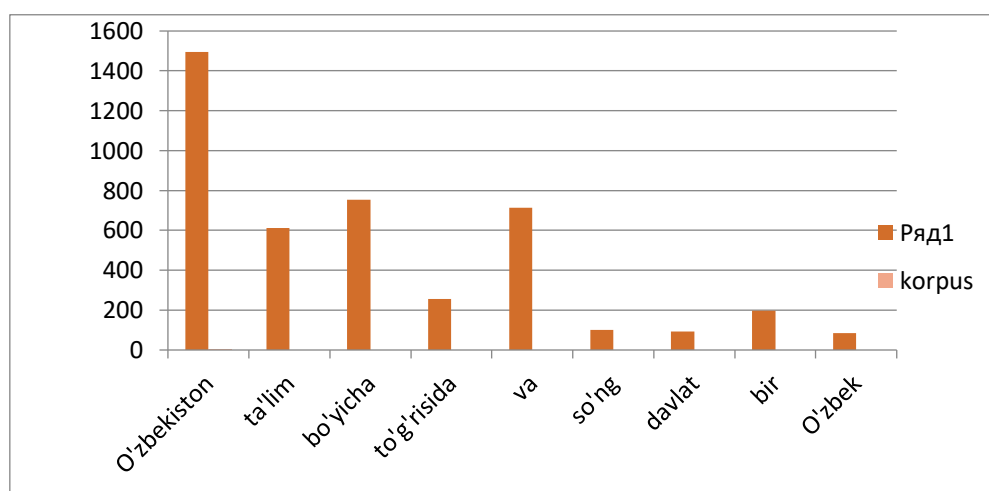
linguists see corpus linguistics as a research tool, while other linguists consider it as a separate science. The goal of developing a newspaper text corpus is to create a machine-readable collection of Uzbek newspaper texts. The collection allows users to study the language of newspapers of a certain period and compare the language of different newspapers within the period. Newspaper corpora are a rich source of neologism and other lexicographic phenomena. The creation of newspaper corpora makes it possible to summarize thousands of articles in one place, which makes it easier for researches to conduct scientific research; it takes less time and less effort. One of the main and major sources of newspaper texts is Common Language Resource and Technology Infrastructure, CLARIN. The CLARIN Eric infrastructure provides access to 35 newspaper corpora, 7 of which are multilingual and 28 monolingual corpora. This corpus contains newspaper articles in 11 languages; Arabic, Czech, Finnish, French, German, Greek, Italian, Norwegian, Polish, and Swedish. In addition, the European collection section of the CLARIN infrastructure gives access to all historical newspapers of Europe. The collection is divided into 9 sub-collection by country. Here we have the opportunity to review individual issues of the newspaper. Another effective newspaper corpus of newspaper texts is the corpus of English newspapers, namely SIBOL. This corpus consists of articles collected from various English language newspapers between 1993-2021. The SIBOL corpus consists of 850 million words in 2 million articles in 18 newspapers. One of the main reason for creating a corpus of newspaper texts is that it clearly shows the processes of language change. The year of publication of the articles included in the corpus is clearly indicated, which serves as a very useful and convenient resource for scientific researchers and seekers. Newspaper texts in Uzbekistan as the first stage of creation, under the leadership of Nilufar Abdurakhmonova, we created the electronic corpus inserted by example of «Ma'rifat» newspaper. The purpose of creating a corpus of newspaper texts is to collect all newspaper articles and texts in one place. This corpus serves as a study object of various fields.

Discussion and Results.

Changes in the field of education and other similar fields can be studied through this corpus. Each article is entered in the corpus in the year of publication. When creating the corpus of the newspaper, first, articles from all issues of the «Ma'rifat» newspaper within certain years were collected, and this article were uploaded to the newspaper corpus section of the Uzbek corpus.



By creating a newspaper corpus, it was possible to analyze the results by field. In the corpus, it was possible to find articles related to economics, politics, law, education, and a number of similar fields in one place, and a number of terms related to these fields were also created. This corpus serves as a primary reference for academic researchers and professors in the field. In the corpus of the created newspaper, the word «Uzbekistan» appeared 1496 times, the word «education» 611 times, the word «by» 754 times, the word «about» 257 times, the word «Uzbek» 190 times, and «initiator» 714 times, «one» participated 196 times.



Conclusion.

In conclusion, it can be said that the creation of a corpus of newspaper texts in Uzbek corpus linguistics makes a small contribution to the development of NLP[natural language processing]. In particular, this corpus allows researchers to study the level of usage and methodological aspects of words in newspapers in a certain period of time. The most convenient and useful part is that brings together thousands of articles by style and period. Creating a corpus of newspaper texts will contribute to the development and expansion of the science of computer linguistics. In the future, we will do a lot of work to expand the electronic corpus and use this corpus to analyze their language, and based on these, we will create a frequency dictionary of concordances.

References:

1. Abdurakhmonova N, Arabboyeva G. XRONOLOGIK LINGVISTIK KORPUS YARATISHDA GAZETA MATNLARI O'RGANISH OBYEKTI SIFATIDA. Academicresearchineducationalsciences, 2022.
Havola: <https://cyberleninka.ru/article/n/xronologik-lingvistik-korpus-yaratishda-gazeta-matnlari-o-rganish-obyekti-sifatida>
2. An-Nahar Newspaper Text Corpus, ELRA, 2010. [Online].
Havola: <http://catalog.elra.info/en-us/repository/browse/ELRA-W0027/>
3. M. Baroni, S. Bernardini. Introducing the La Repubblica Corpus; A Large, Annotated, TEI[XML]-compliant Corpus of Newspaper Italian.
Havola: https://www.academia.edu/download/35013473/rep_lrec_2004.pdf

4. MLCC Multilingual and Parallel Corpora, ELRA, 2012. [Online].
Havola:[<http://catalog.elra.info/en-us/repository/browse/ELRA-W0023/>]
5. Nordquist R. Definition and Examples of Corpus Linguistics, ThoughtCo, 2019.[Online].
Havola:[<https://www.thoughtco.com/what-is-corpus-linguistics-1689936>]
6. Tina Krenmayr. What corpus linguistics can tell us about Metaphor use in newspaper texts. 2014. [Online].
Havola:[https://www.academia.edu/download/35013473/rep_lrec_2004.pdf]

INGLIZ VA O'ZBEK TILI FRAZEMALARIDA INSON QIYOFASI VA XULQ-
ATVORI IFODASI**Yandashova Tursunoy Rustam qizi**

O'zbekiston Milliy universiteti,

O'zbek filologiyasi fakulteti

O'zbek tilshunosligi kafedrasida dotsenti v.b. PhD.

ORCID 0000-0002-5053-1083.

E-mail: yandashova92@mail.ru

Abstract: This article is available in English and Uzbek, representing the concept of Beauty the semantic-structural research of phraseologisms was carried out. The peculiarities and commonalities belonging to the cultures of the two nations are noted separately. Phraseological units are important as units that make our language effective, meaningful and alive. In linguistics, phraseology has been studied as part of lexicology for many years. Because phraseological units are the equivalents of words in the language, and lexicology is the field that studies the words that make up the vocabulary of the language and their equivalents. In the following years, phraseology was separated from lexicology as a separate branch of language or an independent language stage from the 50s of the 20th century. Phraseologisms are the wealth of the nation, and its sources are closely related to the nature, economic system, history, culture, lifestyle, oral creativity, fiction, art, science, and customs of the place where this or that nation lives. The fact that idioms are the most vivid culture-carrying units in linguistics is proven once again by the example of the following ideas.

Keywords: phraseologism, lexicology, culture, reality, ethno-cultural unity, lively communication environment, international phraseology, universal content, uniqueness.

Аннотация: Данная статья доступна на английском и узбекском языках, представляющих концепт Красоты, проведено семантико-структурное исследование фразеологизмов. Отдельно отмечены особенности и общие черты культур двух народов. Фразеологические единицы важны как единицы, которые делают наш язык эффективным, осмысленным и живым. В лингвистике фразеология уже много лет изучается как часть лексикологии. Потому что фразеологизмы – это эквиваленты слов в языке, а лексикология – это область, изучающая слова, составляющие словарный запас языка, и их эквиваленты. В последующие годы фразеология выделилась из лексикологии как отдельная отрасль языка или самостоятельный языковой этап с 50-х годов 20 века. Фразеологизмы являются достоянием нации, а их истоки тесно связаны с природой, хозяйственным строем, историей, культурой, бытом, устным творчеством, художественной литературой, искусством, наукой, обычаями места проживания того или иного народа. Факт То, что фразеологизмы являются наиболее яркими культуруносными единицами в языкознании, еще раз доказывается на примере следующих идей.

Ключевые слова: фразеологизм, лексикология, культура, реальность, этнокультурное единство, живая среда общения, интернациональная фразеология, универсальное содержание, уникальность.

Annotatsiya: Ushbu maqolada o'zbek va ingliz tillaridagi Go'zallik konseptini ifodalovchi frazeologizmlarning semantik-struktur tadqiqi amalga oshirilgan. Bunda ikki millat madaniyatiga tegishli o'ziga xosliklar va umumiyliklar alohida qayd etilgan. Ma'lumki, frazeologik birliklar tilning jonli, ta'sirchan va mazmundor bo'lishida muhim ahamiyatga ega. Frazeologiya uzoq yillar mobaynida tilshunoslikda leksikologiyaning tarkibiy qismi sifatida o'rganib kelindi. Boisi frazemalar tilda so'zlarga ekvivalent hodisa sifatida mavjud bo'lib, so'zga funksional teng hodisalarni leksikologiya tadqiq etardi. So'nggi yillarda frazeologiya tilshunoslikning alohida yo'nalishi sifatida, XX asrning 50-yillaridan boshlab mustaqil bir bo'lim sifatida ajralib chiqdi. Frazeologizmlar millat boyligi sifatida xalqning iqtisodiy hayoti, tarixi, madaniyati, turmush tarzi, ijodkorligi, san'ati, ilm-fan va urf-odatlar bilan bevosita chambarchas bog'liq hisoblanadi. Frazeologizmlarning haqiqatdan ham madaniyat tashuvchi birliklar ekanligi haqidagi fikrlarimizni maqolada keltirilgan namunalar orqali asoslaymiz.

Kalit so'zlar: frazeologizm, leksikologiya, madaniyat, etnomadaniy birlik, jonli muloqot muhiti, internatsional frazeologizm, universal mazmun, o'ziga xoslik.

Frazeologik birlik til hodisasi sifatida lison va nutqqa daxldor birlikdir. Birdan ortiq mustaqil leksema ko'rinishining birikuvidan tashkil topib, obrazli ma'noviy tabiatga ega bo'lgan lisoniy birlik frazeologik birlik deyiladi: *tepa sochi tikka bo'ldi, sirkasi suv ko'tarmaydi; to show the white feather, to play the first fiddle*. Frazeologik birlik ibora, frazeologizm, turg'un birikma, barqaror birikma, frazeologik birikma, idioma atamaları bilan ham yuritiladi.

Frazeologik birlik tashkil etuvchisiga ko'ra qo'shma leksema, so'z birikmasi va gapga o'xshaydi. Biroq ular ko'proq qo'shma leksema kabi til jamiyati ongida tayyor va barqaror holda yashaydi. Boshqacha aytganda, frazeologik birlik lisonda barcha lisoniy birliklarga xos bo'lgan umumiylik tabiatiga ega va nutqda xususiylik sifatida namoyon bo'ladi. Demak, frazeologik birikma nutq hodisasi emas, balki til hodisasidir. Shuningdek, frazeologik birlik nominativ ma'no anglatadi, shunga ko'ra leksema bilan bir qatarga qo'yiladi, leksemadan yirik nominativ birlik deb yuritiladi.

O'zbek frazeologiyasining izchil ilmiy asosda tadqiq qilinishi Shavkat Rahmatullayev nomi bilan bog'liqdir. Olim bu sohaga bag'ishlangan bir qator tadqiqotlari bilan milliy frazeologiyamizning yaratilishiga asos soldi. Jumladan, olimning quyidagi tadqiqot ishlari ya'ni "O'zbek tilining izohli frazeologik lug'ati", "O'zbek tilining qisqacha frazeologik lug'ati", "O'zbek frazeologiyasining ba'zi masalalari", "Frazeologik birikmalarning asosiy ma'no turlari" [Rahmatullayev, 1978: 5] kabi ilmiy adabiyotlari bilan o'zbek frazeologiya maktabi rivojiga o'z hissasini qo'shdi.

O'zbek frazeologiyasining yangi muvaffaqiyatlarga erishuvida B. Yo'ldoshev [Yo'ldoshev, 2007: 88], Abdumurod va Abdug'ofur Mamatovlarning xizmatlari beqiyos hisoblanadi. B. Yo'ldoshev [Yo'ldoshev, 2007: 88] frazeologizmlarning konnotativ ma'nolari, badiiy tasvir vositasi sifatidagi imkoniyatlarini atroflicha yoritib berdi. Abdumurod Mamatov frazeologik norma muammosini ko'tarib chiqdi va frazeologik norma hamda frazeologik variant o'rtasidagi munosabatni ko'rsatib berdi. Abdug'ofir Mamatov esa, o'zining qator asarlarida frazeologik shakllanish muammolarini o'rta qo'ydi va uning paydo bo'lish omillarini yoritib berdi. Frazeologik birliklarning muhim farqlovchi belgisi mazmuniy qayta shakllanish ekanligini ko'rsatdi. Frazeologik birliklarni ilmiy tadqiq qilishda tilshunoslikda keng tarqalgan distributiv metoddan foydalanish mumkin. Frazeologik birliklarning nutqda ishlatilish xususiyatlarini distributiv metod yordamida o'rganilish orqali ularning struktural-

semantik belgilari chuqurroq aniqlanadi. Undan tashqari frazeologik birliklarni o'rganishda ularning semantik va struktural belgilarini tilning mazmun plani bilan uning ifoda plani munosabati borasida qaramoq zarur. Frazeologik birliklarning tarkibini tekshirishda uning komponentlari o'rtasidagi birikish turlarini, birikuvchi so'zlarning tabiatini valentlik metodi asosida o'rganish mumkin.

Yuqoridagilardan kelib chiqqan holda frazeologizmlarni biz chinakam etnomadaniy birliklar deb atashimiz mumkin. Bunday etnomadaniy birliklarni o'rganish va shu asosda muayyan xalq madaniyati, mentalitetidan xabardor bo'lish kabi masalalar lingvokulturologiyada frazeologizmlarni o'rganishdan kelib chiquvchi bosh maqsadlarimizdan biridir. Ayniqsa, frazeologizmlarni qiyosiy aspektda o'rganish asosida xalqlar, mentalitetlar o'rtasidagi o'ziga xosliklar yoki aksincha, mushtarakliklarni aniqlash oson kechadi. Bugungi kun lingvokulturologiyasida ayni shu aspektdagi ishlar katta miqdorni tashkil etadi.

Ingliz tilida tashqi ko'rinish uyg'unligi, go'zallik parametrlari haqida juda ko'p obrazli frazeologik birliklar mavjud. Quyida ularning bir nechtasini tahlilga tortamiz:

- *thin on the top* (tepasi siyrak): sochlari to'kilib, kal bo'lib boshlamoq, juda siyraklashmoq. *Dad's gone a bit thin on the top in the last few years.* O'zbek tilida ham shunga o'xshash iboralar mavjud. Masalan, *tepakal bo'lib qolmoq* yoki *boshidan nur taralmoq*. Ikkinchi ibora ko'proq so'zlashuv nutqida ishlatiladi, adabiy til uchun xos emas.

- *(as) ugly as sin* (juda xunuk): juda xunuk hodisa-voqea, predmet, hayvon yoki shaxslarga nisbatan ishlatiladi. *Have you seen the new neighbour's dog? It is as ugly as sin!* Bizda bu ibora bir so'zga teng *badbashara* leksemasi yoki *ajinasifat* metaforasi bilan ifodalanishi mumkin.

- *vertically challenged* (vertikaliga yetishmaydi): yumoristik ruhdagi ushbu frazema past bo'yi odamlarga nisbatan ishlatiladi. *High shelves are difficult for vertically challenged shoppers.* O'zbek lingvomadaniyatida bo'yi haddan tashqari uzun odamlarga nisbatan *mirzaterakday bo'lmoq* iborasi qo'llaniladi.

- *mutton dressed as lamb* (qo'y qo'zichoqday kiyingan): o'rta yosh yoki undan katta yoshdagi ayollarning yosh ko'rinish maqsadida yosh qiz-ayollar uchun mo'ljallangan kiyimni kiyishi. *The style doesn't suit her – it has a mutton-dressed-as-lamb effect on her!* O'zbek tilida *Kampirning dardi g'ozada* maqoli ushbu iboraga qisman mos keladi.

- *pretty as a picture* (suratday go'zal): juda go'zal va jozibali ko'rinishga ega shaxslarga nisbatan ishlatiladi. *The young bride looked as pretty as a picture in her beautiful dress.* O'zbek milliy madaniyatida bu iborani *qo'g'irchoqday go'zal* obrazli ifodasiga qiyos qilish mumkin.

- *(all) skin and bone* (et bilan suyak): ushbu frazema juda ozg'in kishilarga nisbatan ishlatiladi. *After trekking in the Himalayas, he was all skin and bone.* Har ikki til madaniyatida ham juda ozg'inlik xush qabul qilinmaydi, va bu til egalarining frazeologik fondidan ham munosib joy egallagan: *eti suyagiga yopishgan*; *eti ustixoniga yopishgan*, *cho'pday ozg'in* va b.

- *not a hair out of place* (o'rnidan boshqa joyda bir tola soch yo'q): ko'rinishi mukammal, tashqi qiyofasida, liboslarida hech bir kamchiligi yo'q kishilarga nisbatan ishlatiladi. *Angela is always impeccably dressed – never a hair out of place!* Inglizlarda kiyinish madaniyati juda yuqori baholanadi. Buni yuqoridagi frazema misolida ham ko'rish mumkin. O'zbek lingvomadaniyatida esa *yal-yal yongan*, *yulduzday charaqlagan* iboralari vositasida tashqi ko'rinish mukammalligi ifoda qilinadi.

- *look a sight* (bir ko'rishlik): ko'rinishi juda yomon, kiyimlari betartib shaxslarga nisbatan qo'llaniladi. *She looks a sight in that dress! Pati yulingan tovuqday* iborasi qisman mos kelishi mumkin, biroq bu iboraning mazmunini to'liq ochib berolmaydi.

- *look like a million dollars* (million dollarga o'xshab): ko'rinishi juda yaxshi insonlarga nisbatan ishlatiladi. *With a tan and a new hairstyle she looked like a million dollars!* O'zbek tilida *Quling o'rgilsin!*, *Joningdan*, *akang*, *Aylanib ketay*, *o'rgilib ketay* kabi iboralar aynan yaxshi ko'rinishli odamlarga nisbatan ishlatiladi, biroq mazkur iboraga to'la mos emas.

- *Face only a mother could love* (faqat onasi yaxshi ko'ra oladigan yuz): juda xunuk odamlarga nisbatan kulgili usulda ishlatiladigan frazema. *The poor guy has a face only a mother could love.*

- *face that would stop a clock* (soatni to'xtatib qo'yadigan yuz): bu ham juda xunuk odamlarga nisbatan kulgi ifodalash maqsadida qo'llaniladigan ibora. *You'll recognize him – he's tall and thin, with a face that would stop a clock!* O'zbek xalqida *so'xtasi sovuq*, *turqi sovuq* iboralari shu mazmunda ishlatiladi.

- *five o'clock shadow* (Soat besh soyasi): oxirgi bir kun davomida soqoli olinmagan erkaklarning yuziga nisbatan ishlatiladigan ibora. *He looked tired and had a five o'clock shadow.* *Mo'ylovi sabza urgan* iborasi qisman mos kelishi mumkin, biroq bu iboraning mazmunini to'liq ochib berolmaydi.

- *face like a bulldog chewing a wasp* (ari chaynab qo'ygan baldog yuziga o'xshab): odamlarga nisbatan yuzida juda xunuk ifoda ko'rinish qolganda ishlatiladigan ibora. *Not only was he rude but he had a face like a bulldog chewing a wasp.* *Dimog'iga qurt tushgan*, *qovog'idan qor yog'adi*, *peshonasi tirishgan* iboralari aynan shu holatni ifodalashga xizmat qiladi.

- *down at heel* (tovonidan pastda): puli yo'qligi tufayli juda g'arib, betartib kiyingan odamlarga nisbatan ishlatiladigan ibora. *The down-at-heel student I first met became a successful writer.* *G'aribi bechora*, *Almisoqdan qolgan* iboralari juda yomon kiyingan odamlarga, ularning liboslariga qarab ishlatiladi.

- *dressed up to nines* (to'qqizlargacha kiyingan): juda zamonaviy va qulay kiyingan odamlarga nisbatan ishlatiladi. *Caroline must be going to a party – she's dressed up to nines.* O'zbek tili egalari nutqida qo'llanuvchi *o'ta ko'zga tashlanmoq*, *mingtani ichida bitta bo'lmoq*, *ring bo'lmoq* (so'zlashuv nutqida) iboralari o'ziga juda zeb bergan, zo'r kiyingan shaxslarga nisbatan ishlatiladi.

- *dressed to kill* (o'ldirish uchun kiyingan): o'ta zamonaviy, eng so'nggi urfdagi kiyimni kiymoq. *She arrived at the reception dressed to kill.* O'zbek lingvomadaniyatida *zamondan oldinda yurmoq*, *zamonni ortda qoldirmoq* iboralari kiyinishga nisbatan ishlatilganda xuddi shu ma'noda qo'llaniladi.

- *cut a dash* (chiziqni kesish): chiroyli libos va jozibali ko'rinish bilan odamlarda chuqur taassurot qoldirmoq. *Wearing his uniform, my grandfather cut a dash on his wedding day.* *Ochilib ketmoq*, *chiroyi ochilmoq* iboralari o'zbek madaniyatida shu ma'noda ishlatiladigan iboralar sanaladi.

- *deck out* (bezanmoq): maxsus kunlar uchun biror shaxsni yoki narsa-buyumni bezatish yoki betakror kiyintirish. *Paul decked out his car for the occasion.* O'zbek lingvomadaniyatida *o'zini shaylamoq*, *otini shaylamoq*, *o'zini chog'lamoq* iboralari aynan shu ma'noda qo'llaniladi.

- *set one's heart on* va *set one's mind on*– biron narsaga intilmoq m'nosini bildiradi. Lekin *set one's heart on* biron narsani yurakdan istamoq, butun qalbi bilan intilmoq ma'nosini bildirsa, *set one's mind on* biron narsani istamoq, xohlamoq ma'nosini anglatadi.

Yuqoridagi inglizcha iboralarning qavs ichida keltirilgan tarjimalari so'zma-so'z tarjima bo'lib, shu millat mentaliteti bilan bog'liq tushunchalarni yanada yorqinroq ifodalash maqsadida tarjima qilindi, shu boisdan ba'zi iboralar tarjimasida mazmuniy g'alizliklarga yo'l qo'yilgan bo'lishi mumkin.

Inglizlarda insonning go'zal ko'rinishi *sweets to the sweet, out of this world, graceful as a swan, eye-appeal, feast one's eyes on* kabi iboralarda ifodalanishini ko'rishimiz mumkin. O'zbek tilida esa *ko'z(i)ga o'tday ko'rinoq; ko'zni qamashtirmoq, ko'zga tashlanmoq; ko'zga yaqin bo'lmoq* kabi iboralarni misol keltirishimiz mumkin. Ingliz milliy madaniyatida *golden hair, hair shone glossily; crisp gold hair*, og'iz va lablar go'zalligi o'zbek milliy madaniyatida *gulning g'unchasiday og'iz* iborasi ingliz milliy madaniyatida *small cherry mouth* sifatida ifodalanadi.

Shuningdek, mazkur millat vakillari inson tana a'zolari go'zalligi orqali *the blue eyed boy; the golden (white) headed boy; the fair-haired* iboralari yordamida o'zbek tilidagi *ko'zining oq-u qorasi, o'mganidan yel o'tmagan, yelkasi shamol ko'rmagan* arzanda farzand yoki boshqa juda aziz bir insonni ifodalaydilar.

Beautiful as lily; as rose in June; as May singari tabiat hodisalari, *like candy, honey, fresh bread, cheese-cake* singari iste'mol mahsulotlarining iboralar orqali inson qiyofasi va psixologiyasiga ko'chishi har ikki til madaniyatida ham birday uchrashini kuzatamiz.

O'zbek lingvomadaniyatida ayollar go'zalligi *oy desa oydek, kun desa kundek* kabi tabiat obrazlari, *suv ichsa tomog'idan ko'ringudek; bir qultum suv bilan yutib yuborgudek* kabi gastronomik, *dutorning toriday* kabi predmetlar nomi bilan ayol go'zalligini ta'riflash mumkin.

Ayol va erkak go'zalligini frazeologik iboralarda ilohiy va afsonaviy timsollar yordamida ham qiyoslab o'rganish mumkin. Masalan, inglizcha *young Greek god, like Aphrodite, Venus, angel*; o'zbekcha *Yusufdek, Alpomishdek, parivash, sanam, malak* kabi misollar orqali farqlanadi. Fikrimizcha, *maftun bo'lmoq, lol bo'lmoq, ko'zni ololmay qolmoq* iboralari ingliz va o'zbek lingvomadaniyatiga xos universalliklardan biridir: *eye appeal, have one's eyed glued on* va hk.

Frazeologik birliklar turfa aspektlar doirasida tasniflanishi mumkin. Mazmun planiga ko'ra, struktur tuzlishiga ko'ra, uslubiy ifodalanishiga ko'ra va hokazo. Xususan, uslubiy xoslanishiga ko'ra o'zbek va ingliz tillarida so'zlashuv uslubi frazeologiyasi faol sanaladi. Bunda so'zlashuv frazeologizmlari xalq og'zaki ijodida, kundalik muloqotda keng qo'llanilib, umuman uy-ro'zg'or buyumlari, kiyim-kechak nomlari, oziq-ovqat nomlari, pul birliklari, kundalik ish-harakatlarni, shuningdek, insonning emotsional-ruhiy holatlarini ifodalab keladi.

Foydalanilgan adabiyotlar:

1. Brewer E.C. Dictionary of Phrase and Fable. – Philadelphia; London: James Brodie, 2003. – 1440 p.
2. Кунин А.В. Англо-русский фразеологический словарь. – Изд. 4-е, переработанное и дополненное. - М.: Русский язык, 1984. – 942 с. (АРФС).
3. Rahmatullayev Sh. O'zbek tilining izohli frazeologik lug'ati. T., 1978.
4. Йулдошев Б. Ўзбек фразеологияси ва фразеографиясининг шаклланиши ҳамда тараққиёти. - Самарқанд: СамДУ, 2007.- 108 б.
5. Kurbanova M, Yuldashev M. 2014. Matn tilshunosligi. Toshkent, 154 b.

O'ZBEK STEMASI ALGORITMI

Ismailov Alisher Shakirovich,
Qodirova Gulnoza To'xtasinjon qizi,
Yigitaliyeva Maftuna Aliasqar qizi,
Xusanova Rayxona.
Toshkent Moliya instituti

Annotatsiya: Ushbu maqolada an'anaviy qoidalarga asoslangan tizimning simli o'xshashlik yondashuvidan iborat yangi kontekstsiz bepul stemmer taklif etiladi. Ushbu algoritmi gibriddan iborat deb atash mumkin. Bu tilga bog'liq algoritmi. Kontekstdagi bepul stemmer degani, kontekstga asoslanmagan so'zni ishlatadigan stemmer, degan ma'noni anglatadi, har bir kontekst uchun bunday qoida qo'llaniladi. An'anaviy kontekst bo'yicha erkin qoidaga asoslangan yondashuv yordamida so'zlarni olib tashlaganimizdan so'ng, u dinamik dasturlashning simli o'xshashlik funksiyasini qo'llash orqali yengib o'tilgan so'zlarning tagiga yoki ustiga tushishi mumkin. Satrlarning o'xshashligi funksiyasini o'lchash uchun tahrirlash masofasidan foydalaniladi. O'chirilgan inflatsiya qilingan so'z matn bazasida mavjud bo'lgan so'zlar bilan taqqoslanadi. Minimal masofaga ega bo'lgan so'z, olib tashlangan inflatsiya qilingan so'zning o'rnini egallaydi, bu esa uning kelib chiqishiga olib keladi. Ushbu yondashuvda an'anaviy qoidalarga asoslangan tizim va korpusga asoslangan yondashuv tushunchalari keng qo'llaniladi. Ushbu algoritmi o'zbek tili uchun sinovdan o'tkaziladi.

Kalit so'zlar: Stemming, Text-mining, algoritmi, o'zbek tili.

Abstract: In this paper, a new context free stemmer is proposed which consists of the combination of traditional rule based system with string similarity approach. This algorithm can be called as hybrid algorithm. It is language dependent algorithm. Context free stemmer means that stemmer which stems the word that is not based on the context, for every context such rule is applied. After stripping the words using traditional context free rule based approach, it may over stem or under stem the inflected words which are overcome by applying string similarity function of dynamic programming. For measuring the string similarity function, edit distance will be used. The stripped inflected word is compared with the words stored in a text database available. That word having minimum distance is taken as the substitution of the stripped inflected word which leads to the stem of it. The concept of traditional rule based system and corpus based approach is heavily used in this approach. This algorithm will be tested for Uzbek Language.

Keywords: Stemming, Text-mining, Algorithm, Uzbek language.

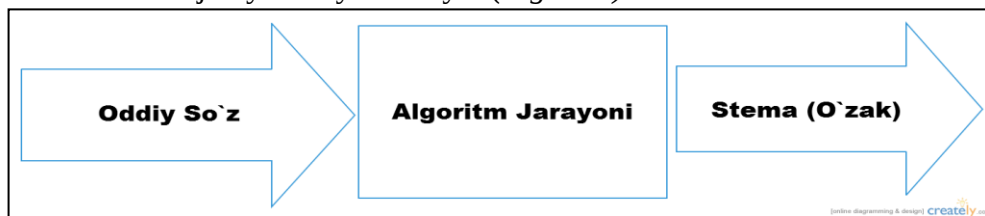
Аннотация: Этот документ предлагает новый свободный от контекста стеммер, состоящий из подхода проводной аналогии системы, основанной на традиционных правилах. Этот алгоритм можно назвать гибридным алгоритмом. Это зависящий от языка алгоритм. Свободный терминатор в контексте означает терминатор, который использует неконтекстное слово, и такое правило применяется к каждому контексту. После того как мы удалили слова, используя свободный подход на основе правил в традиционном контексте, его можно отбросить или свести к словам, которые были преодолены с помощью

функции подобия проводов динамического программирования. Расстояние редактирования используется для измерения функции сходства линий. Удаленное завышенное слово сравнивается со словами, присутствующими в текстовой базе данных. Слово с минимальным расстоянием заменяет удаленное завышенное слово, что приводит к его происхождению. В этом подходе широко используются понятия традиционной системы, основанной на правилах, и корпусного подхода. Этот алгоритм проверен на узбекском языке.

Ключевые слова: стемминг, майнинг текста, алгоритм, узбекский язык.

Stema yoki stemming soʻzni oʻzagini topish deyiladi. Stemming odatda tabiiy til jaroyoni (Natural Language processing), axborot qidiruv (Information Retrieval), (Text Mining) da foydalaniladi.

Misol uchun, “insonlar” soʻzini tabiiy til jaroyoni (Natural Language processing) qidirsak, qidiruv tizimni ishlash jaroyonini sekinlashtirishi mumkin, agar soʻzni oʻzagini ajratib olib qidirsak, tizimni ishlash jaroyoni tezlashadi. Stemming soʻz jamlanmalari (text mining) da juda muhim oʻrin tutadi. Stemming odatda kiritilgan soʻzning oʻzagini ajratib oladi va bu jarayon axborot qidiruvni ishlash jarayonini yaxshilaydi (Figure 1).



1. *Muammo.* Bir stemmerni boshqasidan farqi uni soʻzni qanday oʻzagini ajratib olishda. Buning uchun har xil usullar qoʻllaniladi. Soʻzning oʻzagini topishda prefiks va suffikslarni oʻchirib tashlanadi. Bu usul agressiv stemmer deb nomlanadi. Bu jarayonda bir nechta muammolar mavjud:

- a) soʻzning oʻzagini topishda natija kiritilgan soʻzga bogʻliq boʻlmagan boʻlishi mumkin (over-stemming).
- b) soʻzni prefiks va suffikslardan ajratib olishda soʻz oʻzagiga tegishli boʻlgan harfni ham oʻchirib tashlanishi mumkin (under stemming).

Ushbu muammolar algoritmnini yangi dasturlarda qoʻllanganda qiyinchiliklar tugʻdirishi mumkin. Oʻzbek stema alrotimni rivojlantirishdagi asosiy maqsad yuqorida keltirilgan muammolardan holi algoritim yaratishdir. Ilmiy tilda soʻzni oʻzagini topish lemmatizatsiya deb ataladi [25, 26, 27].

2. Oʻrganilgan adabiyotlar

Stema mavzusida bir qancha ilmiy ishlar mavjud. Misol uchun, nemis, ispan, hind stemalarini keltirish mumkin. Hind stemasi bir nechta hind tillarida ishlaydi. Ular Tamil, Punjabi, Bengali, Gujarati, Hind va marati tillari. Shunga oʻxshash arab stemalari ham mavjud. Har bir stema yaratuvchilarni oʻzlarini tillarida ishlaydi.

[1] stemasi – arab tili stemasi hisoblanadi. Bu stema 5 usuldan foydalanib yaratilgan. Ulardan 4 tasi soʻzdagi harflarni joylashishi usuli va beshinchi usul qoidalarga asoslangan tizimi (traditional rule based system). Arab stemasida qoidalarga asoslangan tuzatish algoritmi bilan birga ishlaganda yaxshi natija bergan.

[2] stema ham arab tili stemasi. Bu stema mashina oʻrganish tizimidan foydalanilgan va

juda yaxshi va aniq natija bergan. Bunday yaxshi natijaga erishish uchun ular Baye va ZeroR va algoritmlaridan foydalanganlar.

[3] - stema yengil stema va axborot qidiruv usullaridan foydalanib yaratilgan.

[4] – stema bir nechta stemming algoritmlari va solishtirish usullaridan foydalanilgan. Ushbu stemada Lovins va Poster stemalaridan foydalanilgan.

[5] – bu maqola stema algoritmini Ingliz tili uchun tayyorlangan va bu stema gibrid usulidan foydalanilgan. Gujarati (hind) va Arab algoritmlaridan foydalanilgan.

3. Taklif qilinayotgan Stema Tizimi (O'zbek stema)

Taklif etilayotgan modelda quyidagilarni hisobga oladi: Prefiks, suffiks va o'zak.

3.1. Suffiks

Suffiks deb so'zni o'zagidan keyingi qo'chimchalarga aytiladi.

3.2. Prefiks

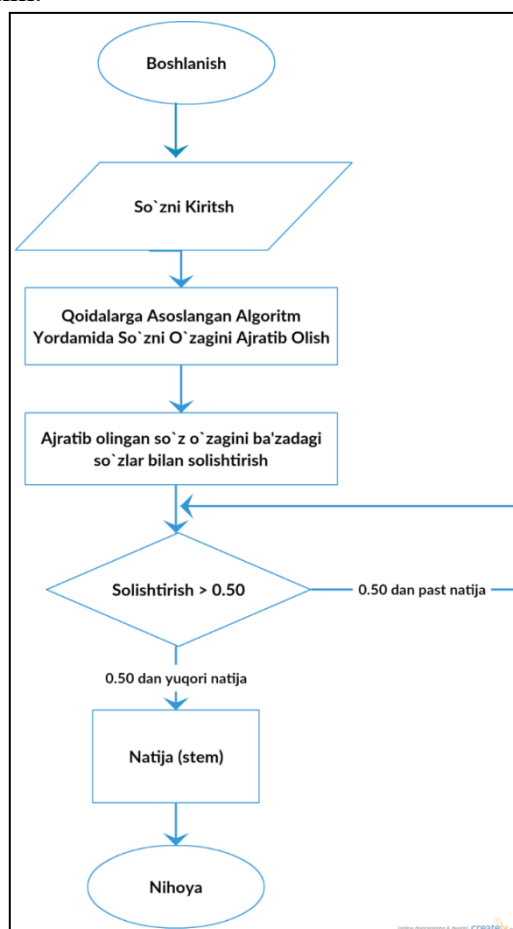
Bu o'zbek so'zlarning oldingi qismidagi qo'shimchalar.

3.3. O'zak so'zlar

Ushbu algoritm uchun 700 ta o'zak so'zlardan foydalaniladi.

3.4. Gibrid Metodologi

Bu algoritm klassik qoidalarga asoslangan tizim va solishtirish tizimlaridan foydalanib yaratiladi. 2-rasmda ko'rish mumkin.



[24] – maqola 0.50 foiz aniqlikda ishlagani uchun biz ushbu maqolani eng yaxshi natija sifatida oldik. So'zni o'zagini ajratib olingandan so'ng uni bazadagi o'zak so'zlar bilan

solishtirish yo'li bilan topiladi. So'zning o'zagini ajratib olishda ba'zida o'zakdan ortiqcha harf olinib ketishi yoki o'zakda ortiqcha harf qolib ketish ehtimoli bor. Shuning uchun o'xshash so'zlar solishtirish(string similarity approach)usuli bilan algoritmda qo'llaniladi.

Foydalanilgan adabiyotlar:

1. Ismailov, A., Jalil, M. A., Abdullah, Z., & Abd Rahim, N. H. (2016, August). A comparative study of stemming algorithms for use with the Uzbek language. In *2016 3rd International conference on computer and information sciences (ICCOINS)* (pp. 7-12). IEEE.
2. Jalil, M. M., Ismailov, A., Abd Rahim, N. H., & Abdullah, Z. (2017). The Development of the Uzbek Stemming Algorithm. *Advanced Science Letters*, 23(5), 4171-4174.
3. Allworth, E. A. (2013). *The modern Uzbeks: from the fourteenth century to the present: a cultural history*. Hoover Press.
4. Dickens, M. (2002). Introduction to the Uzbek Language. Retrieved from oxuscom.com: <http://www.oxuscom.com>
5. Kim, S. B., Seo, H. C., & Rim, H. C. (2004, July). Information retrieval using word senses: root sense tagging approach. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 258-265). ACM.
6. 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;5(03):89-100.
7. N. Mahmudov, A. N. (2007). *Ona tili*. Toshkent: Ma'naviyat.
8. Schlyter, B. N. (2001). Language policies in present-day Central Asia. *International Journal on Multicultural Societies*, 3(2), 127-136.
9. N. Mahmudov, A. N. (2009). *Ona Til1*. Toshkent: Tasvir.
10. Matlatipov, G., & Vetulani, Z. (2009). Representation of uzbek morphology in prolog. In *Aspects of Natural Language Processing* (pp. 83-110). Springer Berlin Heidelberg.
11. Abduraxmonova, N. Z. «Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref.» (2018).
12. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. *Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta*. 2016;2 (38):12-7.
13. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
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, doi: 10.1109/ICISCT52966.2021.9670043.
15. Abdurakhmonova, N., Razakhova B.: Ontology of grammar rules as example of noun of Uzbek and Kazakh languages. In: Abstract of the VI International Conference "Modern

- Problems of Applied Mathematics and Information Technology - Al-Khorezmiy 2018”, pp. 37–38, Tashkent, Uzbekistan (2018)
16. Abduraxmonova, N. Z. «Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aforef.» (2018).
 17. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
 18. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
 19. 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.
 20. 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;5(03):89-100.
 21. 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
 22. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. In Proceedings of the International Conference on Language Technologies for All (LT4All) 2019.
 23. 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, doi: 10.1109/ICISCT52966.2021.9670043.
 24. Ismailov, A. S., Shamsiyeva, G., & Abdurakhmonova, N. (2021). Statistical machine translation proposal for Uzbek to English. *Science and Education*, 2(12), 212-219.
 25. Mirzakhlov, J., Babu, A., Ataman, D., Kariev, S., Tyers, F., Abduraufov, O., ... & Chellappan, S. (2021). A Large-Scale Study of Machine Translation in the Turkic Languages. arXiv preprint arXiv:2109.04593.
 26. Johanson, L., & Johanson, É. Á. C. (2015). The Turkic Languages. Routledge.

KORPUS LINGVISTIK TA'MINOTINI YARATISH MASALALARI

Eshmamatova Mahliyo Xamzayevna

O'zbekiston milliy universiteti

Kompyuter lingvistikasi

II bosqich magistratura talabasi

m.eshmamatova21@gmail.com

Annotatsiya: Korpus lingvistikasining hozirgi kundagi rivojlanishi va oldiga qo'yilgan masalalarni bilib olishimiz mumkin. Muayyan ma'lumot uchun til ma'lumotlarining vakillik massivi davri tilning leksik tarkibidagi o'zgarish jarayonlari dinamikasini o'rganish, turli janrlarda va turli mualliflar tomonidan leksik va grammatik xususiyatlarni tahlil qilish imkonini beradi. Aytish mumkinki, korpus lingvistikasi o'zining predmeti sifatida keng foydalanuvchilar manfaatlari yo'lida lingvistik tadqiqotlar uchun mo'ljallangan lingvistik ma'lumotlarning reprezentativ massivlarini yaratish va ulardan foydalanishning nazariy asoslari va amaliy mexanizmlariga ega.

Kalit so'zlar: korpus, lingvistik ta'minot, korpus lingvistikasi, matnlar korpusi.

Abstract: We can learn about the current development of corpus linguistics and the issues that are set before it. For specific information, the period of the representative array of language data allows to study the dynamics of change processes in the lexical structure of the language, to analyze lexical and grammatical features in different genres and by different authors. It can be said that corpus linguistics as its subject has theoretical foundations and practical mechanisms for creating and using representative arrays of linguistic data intended for linguistic research for the benefit of a wide range of users.

Keywords: corpus, linguistic support, corpus linguistics, corpus of texts.

Korpus lingvistikasi - kompyuter texnik va texnologiyalaridan foydalangan holda lingvistik korpuslarni (matnlar korpuslarini) ishlab chiqish va ulardan unumli foydalanishning umumiy tamoyillarini ishlab chiqadigan hisoblash tilshunosligining bir tarmog'i. Lingvistik yoki matnlar korpusi (yoki odatda shunchaki matnlar korpusi) bir ma'noda lingvistik muammolarni hal qilish uchun tashkil qilingan, birlashtirilgan, mo'ljallangan, belgilangan, filologik jihatdan malakali til ma'lumotlarining mashina o'qiy oladigan formati sifatida tushinish mumkin. Shu sababli, yechish maqsadida u yoki bu korpus yaratilgan vazifalar doirasini (ba'zan ancha keng) hisobga olsak, korpus har doim pragmatik yo'nalishni yo'lga qo'yishimizni aytishimiz mumkin.

Hozirgi kunda "korpus" atamasining ko'plab ta'riflari mavjud. Masalan, Э. Финегана darsligida berilgan ta'rifda shunday deyilgan: korpus - bu matnlarning vakillik yig'indisi bo'lib, odatda mashinada o'qiladigan formatda bo'lib, matn yaratilgan vaziyat haqidagi ma'lumotlar, masalan, ma'ruzachi, muallif haqidagi ma'lumotlar kiradi. Vikipediya korpusni statistik tahlil va gipotezalarni tekshirish, til qoidalarini tasdiqlash yoki asoslash uchun foydalaniladigan katta va tuzilgan matnlar to'plami (hozir odatda elektron shaklda) sifatida belgilaydi.

Т. Мак-Энери и Э. Вилсон quyidagi ta'rifni beradi: korpus - bu til modeli sifatida foydalanish uchun aniq til mezonlariga muvofiq tanlangan til qismlari yig'indisidir. Yuqoridagi

ta'riflar zamonaviy matnlar korpusining asosiy xususiyatlarini ta'kidlaydi: maqsad ("mantiqiy g'oya", pragmatik yo'nalish), mashinada o'qiladigan format, matn tanlashning maxsus protsedurasi natijasida reprezentativlik, metallingvistik ma'lumotlarning mavjudligi. Og'zaki materialni mashinada standartlashtirilgan taqdimoti uni qayta ishlash uchun standart dasturlardan foydalanishga imkon beradi. Yaratishning maqsadga muvofiqligi va holatlardan foydalanishning ma'nosi quyidagi shartlar bilan belgilanadi:

- korpusning yetarlicha katta (vakillik) va muvozanatlashgan hajmi ma'lumotlarning tipikligini kafolatlaydi va til hodisalarining butun spektrini taqdim etishning to'liqligini ta'minlaydi;
- har xil turdagi ma'lumotlar korpusda tabiiy kontekstual shaklda bo'lib, ularni har tomonlama va ob'ektiv o'rganish imkoniyatini yaratadi;
- bir marta yaratilgan va tayyorlangan ma'lumotlar massivi turli tadqiqotchilar tomonidan va turli maqsadlarda qayta-qayta ishlatilishi mumkin.

"Matnlar korpusi" tushunchasi boshqaruv tizimini ham o'z ichiga oladi, korpus menejeri (yoki korpus menejeri) deb ataladigan matn va lingvistik ma'lumotlar. Bu korpusda so'ralgan ma'lumotlarni qidirish va foydalanuvchiga qulay shaklda taqdim etish hamda statistik ma'lumotlarni olish uchun dasturiy vositalarni o'z ichiga olgan ixtisoslashtirilgan qidiruv tizimi. Korpusda qidirish har qanday so'z uchun muvofiqlikni yaratish imkonini beradi, manbaga havolalar bilan kontekstda berilgan so'zning barcha qo'llanilishi ro'yxati. Biroq, bundan tashqari, korpusdan matn yoki leksik birliklarning xususiyatlari, til birliklari va lingvistik kategoriyalar va metama'lumotlar (so'z shakllari, leksemalar, grammatik kategoriyalarning chastotasi, turli davrlarda chastotalar va kontekstlarning o'zgarishi) to'g'risidagi statistik ma'lumotlar haqida ma'lumot olish uchun foydalanish mumkin. vaqt), qo'shma hodisa haqidagi ma'lumotlar leksik birliklar, janr va stilistik xususiyatlar va boshqalar. Ushbu statistik ma'lumotlar to'g'ridan-to'g'ri chiqarilishi mumkin (masalan, chastotalar ro'yxati) yoki ular "ichki" hisob-kitoblar va korpusga bevosita kiritilmagan yangi ma'lumotlarni chiqarish uchun ishlatilishi mumkin, masalan, matndagi birikmalar barqarorligini miqdoriy ifodalash, paradigmatic (semantik) klasterlash, leksik birliklar, matnning tayanch so'zlarini aniqlash.

Muayyan ma'lumot uchun til ma'lumotlarining vakillik massivi davri tilning leksik tarkibidagi o'zgarish jarayonlari dinamikasini o'rganish, turli janrlarda va turli mualliflar tomonidan leksik va grammatik xususiyatlarni tahlil qilish imkonini beradi.

Tilshunos tadqiqotchilar funktsiyalarga tobora ko'proq qiziqmoqda qo'shimcha qilib aytish mumkinki, korpus ma'lumotlarini intellektual qayta ishlash va shunday dasturlar mavjudki, ular endi faktografik tipdagi axborot- qidiruv tizimi sifatida korpus menejeri emas, balki turli xil lingvistik funktsiyalarning keng doirasini ta'minlaydigan dasturiy ta'minot, lingvistik, matematik vositalarning murakkab konglomeratidir. Biz ushbu kontsepsiya uchun "korpus xizmati" nomini taklif qilamiz.

Bugungi kunda korpus tilshunosligi ko'pincha kompyuterlar va elektron korpuslar orqali real hayotda tildan foydalanishni o'rganish bilan bog'liq bo'lgan yangi lingvistik intizom sifatida tushuniladi. Korpus lingvistikasida mustaqil fan sifatida asos bo'ladigan kamida ikkita xususiyat mavjud: 1) og'zaki nutqning tabiiy materiallar, ya'ni belgilangan matnlar; 2) asboblarga to'plamining o'ziga xos xususiyatlari. Agar tilshunoslikning sintaksis, semantika va sotsiolingvistika kabi tarmoqlari til tuzilishi yoki til qo'llanilishini tavsiflash yoki baholashga

qaratilgan bo'lsa, korpus lingvistikasi kengroq tushuncha bo'lib, u nafaqat tilshunoslikning, balki tilshunoslikning ham ko'p jihatlariga tatbiq etilishi mumkin.

Korpus usullari asosida “kulturometriya” (kulturomika) deb ataladigan va gumanitar fanlar tadqiqotining barcha sohalariga taalluqli yangi fan. Korpus lingvistikasi “lingvistik tadqiqotlarning turli sohalaridan metodlar to'plami” deb ataladi. Qanaqasiga lingvistik tahlil usuli korpus lingvistikasi, shuningdek, qiyosiy o'rganish jarayonida tillar, lahjalar yoki til variantlari o'rtasidagi umumiy va alohida faktlarni aniqlashga qaratilgan kontrast tadqiqotlar bilan bog'liq. Lingvistik tahlilning ko'p turlari mustahkam va keng empirik dalillar asosida yaxshi rivojlanadi. Umuman olganda, korpus lingvistikasining tilshunoslikdagi o'rni haqida savol tug'ilganda, aftidan, bu amaliyotda qo'llaniladigan lingvistik tadqiqot metodologiyasi desak to'g'ri bo'ladi, tilshunoslikning istalgan sohasiga. Biroq, yana bir qarash bor: korpus lingvistikasi, aslida, haqiqiy ilmiy tilshunoslikdir. Ingliz adabiyotida bu yondashuvlar - korpus lingvistikasi tilshunoslik metodologiyasi va alohida fan sifatida - korpusga asoslangan (korpusga yo'naltirilgan yondashuv) va korpusga asoslangan (korpusga asoslangan yondashuv) deb ataladi. Birinchi yondashuv lingvistik nazariyalarni yoki farazlarni qo'llab-quvvatlash, tasdiqlash, rad etish yoki takomillashtirish uchun korpuslardan foydalanishni nazarda tutadi. Ikkinchi yondashuv korpusning o'zi til haqidagi nazariyalarimizning asosiy va yagona manbai ekanligini e'lon qiladi, korpus lingvistikasi bu yerda nazariya maqomini oladi va ko'rib chiqiladi “tilshunoslik nazariyasidagi eng muhim tushuncha” sifatida. Demak, korpus bilvosita til nazariyasini o'z ichiga oladi va siz faqat u erdan “olishingiz” kerak. Bu tushuncha bizni 20-asrning birinchi uchdan biridagi amerikalik strukturalistlarning ishiga qaytaradi. Korpus lingvistikasining chuqurligida bu yondashuv neofirthian deb ataladi, chunki u J.R.Firth tomonidan kiritilgan kollokatsiya tushunchasi bilan kuchli bog'liqdir. Ehtimol, korpus tilshunosligidagi eng mashhur iqtibosdir

J. R. Firth: “Siz bir so'zni u ushlab turgan kompaniyadan bilib olasiz”. Bu yondashuvning mohiyati shundan iboratki, so'zning ma'nosi (boshqa lingvistik tushunchalar kabi) faqat kontekstda (matnda) mavjud bo'ladi. Ma'lumotlar tahlilchisi hech qanday apriori o'rnatilgan nazariy tushunchalardan foydalanmaydi deb taxmin qilinadi. Til o'rganishga yana bir asosi - bu nutq tushunchasi. Ular uchun nutq faqat matn, tilning “amaliyoti” emas, balki tilning o'zini yoki subtilni amalga oshirish usuli, nafaqat nutq usuli, balki fikrlash tarzidir.

Ushbu yo'nalishda e'tirof etuvchi va korpus resurslari va usullaridan bir qator pozitsiyalarda foydalanadigan olimlarning qarashlari psixolingvistika va sotsiolingvistika bilan qo'shiladi.

V. A. Plungyaning bayonotini ham keltirishingiz mumkin (ma'ruza Sankt-Peterburgdagi Yevropa universitetida) agar ilgari tilshunoslik ikkita “ustun” – lug'at va grammatika ustida turgan bo'lsa, endi ularga uchinchi gipostaz – korpus qo'shildi. Amalda yuqoridagi ikkala yondashuvning umumiy jihatlari ko'p. Va shuni ta'kidlash mumkinki, korpusga asoslangan yondashuvni tan olgan tilshunoslarning ko'plab nashrlari aslida korpusga yo'naltirilgan tadqiqotlardir. Bizning fikrimizcha, korpus lingvistikasiga ta'rif berishning muhim jihati shundaki, u shunchaki tilni o'rganish metodologiyasi emas – o'rganish ob'ekti hosil bo'lgan tizimdir. E. Finegan korpus lingvistikasini korpusni tuzish va ishlatish uchun zarur bo'lgan va tabiiy foydalanishni o'rganishga qaratilgan faoliyat sifatida belgilaydi. Bu ta'rif korpus tilshunosligining ijodiy yo'nalishini ta'kidlaydi. Uning ikki tomonlama tabiati (matn korpusini yaratish va undan foydalanishga qaratilgan) uning ob'ektining ikki tomonlama tabiati - matnlar korpusi bilan belgilanadi, bu esa, bir tomondan, korpus lingvistikasi va boshqa tilshunoslik fanlari uchun manba nutq materialidir;

boshqa tomondan, uning o'zi korpus lingvistikasining mahsulidir. Aytish mumkinki, korpus lingvistikasi o'zining predmeti sifatida keng foydalanuvchilar manfaatlari yo'lida lingvistik tadqiqotlar uchun mo'ljallangan lingvistik ma'lumotlarning reprezentativ massivlarini yaratish va ulardan foydalanishning nazariy asoslari va amaliy mexanizmlariga ega.

Foydalanilgan adabiyotlar:

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;5(03):89-100.
2. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.
3. Abduraxmonova Nilufar. Mashina tarjimasining lingvistik ta'minoti. Monografiya. Toshkent-2018 29- bet
4. Gilmullin, R. (2008). The Tatar-Turkish machine translation based on the two-level morphological analyzer. Interactive Systems and Technologies: The Problems of Human-Computer Interaction, pages 179–186.
5. W.John Hutchins, Machine Translation: A Brief History, Concise history of the language sciences: from the Sumerians to the cognitivists. Edited by E.F.K.Koerner and R.E.Asher. Oxford: Pergamon Press, 1995. Pages 431-445
6. Abduraxmonova N.Z. Mashina tarjimasining lingvistik ta'minoti. Monografiya. Toshkent-2018 30-31- bet
7. Захаров, В. П., & Богданова, С. Ю. (2018). Корпусная лингвистика: учебник для направлений «Лингвистика» и «Педагогическое образование».
8. Mair, C., & Hundt, M. (Eds.). (2000). *Corpus linguistics and linguistic theory* (No. 33). Rodopi.

2-SHO'BA

TIL TA'LIMIDA KOMPYUTER TEXNOLOGIYALARI

O'ZGA TILLARNI O'QITISHDA INTERNETNING O'RNI

Keldiyorova Gulchehra Saydiyevna

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

Annotatsiya: Maqolada jadal kirib borayotgan axborot texnologiyalarini ta'lim jarayoniga ham tatbiq etish dolzarb masalalardan biri ekanligi, kompyuter va axborot texnologiyalaridan foydalanib, ta'lim sohasida, o'quv faoliyatida va o'quvchilar ijodiy tafakkurini rivojlantirishda yangi imkoniyatlar yaratishi, kompyuter vositalari orqali bolaning bilishga oid, axloqiy, ijodiy, muloqot qilish va estetikaviy imkoniyatlarini, qobiliyatlarini ro'yobga chiqarishi, buning uchun esa o'qituvchining o'zi axborot texnologiyalari sohasidan yaxshi xabardor bo'lishi kerakligi, tuzilgan kompyuterli o'qitish dasturlari axborotni o'quvchi tomonidan o'zlashtirilishining psixologik va pedagogik qoidalari e'tiborga olingan taqdirda o'qitishni individuallashtirish va differensiyalashtirishga, bolalarning qiziqishlari va o'qishdagi mustaqil faoliyatini rag'batlantirishga imkon berishi yoritilgan.

Kalit so'zlar: kompyuter, internet- loyiha, interaktiv, kommunikativ kompetensiya, resurs, optimallashtirish.

Аннотация: В статье одним из актуальных вопросов является применение информационных технологий в образовательном процессе, который стремительно входит, используя компьютеры и информационные технологии для создания новых возможностей в сфере образования, воспитательной деятельности и развития творческого мышления обучающихся. , о познании ребенком средствами ЭВМ., реализовать нравственные, творческие, коммуникативные и эстетические возможности и способности, а для этого сам педагог должен хорошо разбираться в области информационных технологий, должна быть освоена информация структурированных компьютерных обучающих программ студентам разъясняется, что учет психолого-педагогических правил позволяет обеспечить индивидуализацию и дифференциацию обучения, стимулирование интересов и самостоятельной деятельности детей в обучении.

Ключевые слова: компьютер, интернет-проект, интерактив, коммуникативная компетенция, ресурс, оптимизация.

Abstract: In the article, it is one of the urgent issues to apply information technologies to the educational process, which is rapidly entering, using computers and information technologies to create new opportunities in the field of education, educational activities and the development of students' creative thinking, about the child's knowledge through computer tools. , to realize moral, creative, communication and aesthetic possibilities and abilities, and for this the teacher himself should be well aware of the field of information technologies, the information of structured computer training programs should be mastered by the student it is explained that if psychological

and pedagogical rules are taken into account, it allows for individualization and differentiation of teaching, stimulation of children's interests and independent activity in learning.

Keywords: computer, internet project, interactive, communicative competence, resource, optimization.

Hozirgi bosqichda jamiyatda sodir bo'layotgan ijtimoiy jarayonlarning o'sishi zamonaviy mutaxassis shaxsini shakllantirishga yangicha yondashuvni belgilaydi. Mas'uliyatni o'z zimmasiga oladigan, tanqidiy va ijodiy fikrlaydigan, boshqa madaniyat vakillari bilan muvaffaqiyatli va samarali hamkorlik qiladigan mutaxassisga talab kuchaymoqda. O'zgaruvchan hayotiy vaziyatlarga moslashish, ma'lumotlar bilan ishlash, mustaqil ravishda kerakli bilimlarni olish, uni turli muammolarni hal qilishda amalda qo'llash qobiliyati zamonaviy inson hayotida tobora muhim o'rin egallab turibdi. Bunday vazifalarni zamonaviy pedagogik strategiyalar va o'qitish texnologiyalarini jalb qilgan holda amalga oshirish kerak. To'plangan ijobiy tajribani saqlab qolish bilan birga ta'lim dasturlarining zamonaviy jamiyat ehtiyojlarini qondirish qobiliyati katta ahamiyatga ega. Internetdagi axborot resurslari va xizmatlari yangi pedagogik texnologiyalarga xos imkoniyatlarni amalga oshirish imkoniyatini yaratib, o'quv jarayoniga prinsipial jihatdan yangi yondashuvni joriy etishga hissa qo'shadi. O'qituvchilarning zamonaviy Internet-resurslarga bo'lgan murojaatlari ta'lim berishda keng ma'lumot olish, izlash va tadqiqot ishlariga jalb qilish asosida o'z bilimlarini rivojlantirish imkoniyatini beradi, bu esa o'z navbatida, til bilimlarini qo'llash va kengaytirish bilan ham bog'liq. Bu mohiyatan insonni boshqa odamlarni bilish orqali o'zini o'zi bilishga, madaniyatlarni, madaniy tafovutlarni va madaniy birlikni, dunyoda mavjud bo'lgan madaniy farqlarga nisbatan bag'rikeng munosabatni o'rganishga qaratilgan ijtimoiy-madaniy ta'limdir. Qidiruv natijalari to'g'risida hisobot berish orqali talaba o'rganilayotgan masala bo'yicha o'z fikrini bildirishi va bahslashishi mumkin. Fikrlarni guruhning boshqa talabalari bilan muhokama qilish mumkin. Ishtirokchilar o'z g'oyalarini asoslashlari, xulosalar chiqarishlari, faraz qila olishlari kerak. Muhokama davomida talabalardan bitta umumiy munozarali savolga javob berishlari so'raladi. Telekommunikatsiyalarning eng keng tarqalgan shakli (ya'ni Internet texnologiyalari orqali aloqa) elektron pochta (elektron pochta), forumlar, chat texnologiyalari, videokonferentsiyalar kabilarni o'z ichiga oladi. Dastlab aloqa uchun yaratilgan, ular bugungi kunda ta'lim maqsadlarida foydalanilmoqda. Bu ta'limni rag'batlantirishning kuchli vositasi, global fikrlash mafkurasi. Telekommunikatsiyalarning eng sodda va eng qulay shakli-elektron pochta-tinglovchilarni chet tilidagi kommunikativ amaliyotga jalb qilishga yordam beradi, yozma tilni takomillashtirishga yordam beradi. Bundan tashqari, ushbu turdagi ishlar og'zaki bayonlarni tuzishda qiynaladigan talabalar uchun tavsiya etiladi, chunki matnni tuzatishga, qayta yozishga imkon beradi. Ta'limning muhim vazifasi zarur ma'lumotlarni qidirish va tanlash qobiliyatini o'rgatish va uni murakkab muammolarni hal qilishda ijodiy tatbiq etishdan iborat ekan, o'quv jarayoniga yangi axborot texnologiyalarini kiritishning afzalliklari aniq. Internet- loyihalar tinglovchilarning ijodiy va tadqiqot faoliyatini rivojlantirishga, yangi ijtimoiy-madaniy bilimlarni shakllantirishga va natijada ijtimoiy-madaniy vakolatlarni shakllantirishga hissa qo'shadi.

Internetning global kompyuter tarmog'i til o'qitishning turli xil modellarini amalga oshirish uchun nihoyatda qimmatli manba, ya'ni tez-tez interaktiv shaklda taqdim etiladigan va amalda mavjud bo'lmagan eng yaxshi ma'lumot manbalaridan biridir. Nofilologik oliy o'quv yurtlarida til o'rgatish haqida gapirganda, uning umumta'lim fanlariga tegishli ekanligini va uni o'rganish muddati cheklanganligini yodda tutish kerak. Biroq jamiyatdagi zamonaviy ijtimoiy-

siyosiy jarayonlar shaxsning lingvistik madaniyatini oshirishni talab qiladi va tillarni amaliy o'zlashtirish zarurligini belgilaydi.

Bularning barchasi bizni ushbu muammoga yangicha qarashga majbur qiladi. Ko'pgina universitetlar xodimlari zamonaviy talablarga muvofiq mutaxassislar tayyorlashga qaratilgan yangi oliy ta'lim paradigmasini amalga oshirmoqdalar. Oliy ta'limning yangi maqsad va vazifalari so'zni keng ma'noda tushuniladigan shaxs madaniyatini rivojlantirish, nutq madaniyati, millatlararo muloqot madaniyatiga javobgarligini anglashni rivojlantirish orqali talaba shaxsining ijodiy salohiyatini ro'yobga chiqarish; o'z-o'zini tarbiyalashga ehtiyojni shakllantirish, axborot olish orqali bilimlarni uzluksiz ravishda to'ldirish va ularni amalda qo'llash zarurligini anglash; o'qitishni interaktiv asosda tashkil etishdan ham iborat. Internetning global kompyuter tarmog'i til o'qitishning turli xil modellarini amalga oshirish uchun juda qimmatli manba hisoblanadi. O'z mohiyatiga ko'ra, internet tez-tez interaktiv formatda taqdim etiladigan va an'anaviy darslikda yoki o'quv qo'llanmasida mavjud bo'lmagan eng yaxshi ma'lumot manbalaridan biridir. Biroq ta'limni internetlashtirish haqida gapirganda, biz Internet didaktik vazifalarni amalga oshirish vositasi ekanligini va lisoniy bo'lmagan mutaxassislik sifatida o'qitishning asosiy maqsadi kommunikativ kompetensiyani, ya'ni qobiliyatni shakllantirish ekanligini anglashimiz kerak. Til o'qitishning o'ziga xos xususiyati shundan iboratki, o'qitishning maqsadi mavzuning o'zi, ya'ni til (lisoniy kompetensiya) haqida bilim emas, balki har xil turdagi ma'lum ko'nikma va malakalarni rivojlantirishdir. Til o'qitishda talabalarning (va har bir talabaning) mustaqil harakatlarini ular o'qitiladigan nutq faoliyati shaklida tashkil etish zarur. Agar o'quvchilar o'qishga o'rgatilsa, unda har bir talabaga o'qish, o'qishga mashq qilish imkoniyati berilishi kerak. Gapirishni o'rgatishda har kimga so'zlash, o'z fikrlarini chet tilida bayon etish imkoniyati berilishi kerak. Tinglashni o'rgatishda har bir talaba o'rganayotgan tilidagi nutqni tinglashi kerak. Taniqli metodist I.V.Raxmanovning fikrini yodda tutish muhimdir: eshitish - harakat qobiliyatlari nutq faoliyatining har qanday turini o'rgatishning asosidir, shuning uchun ko'nikmalarni shakllantirishda nutq faoliyatining har qanday turida og'zaki mashq qilish zarur. Talabalarning til bilish darajalari turlichadir. Shu munosabat bilan, o'qituvchilarda bunday xilma-xil vazifalarga javob beradigan Internet-resurslarni qayerdan topish mumkinligi va dasturning o'zini o'zgartirmasdan, uning samaradorligini oshirish uchun auditoriyadagi ishni to'g'ri tashkil etishi kerak. Jahon tarmog'ida mavjud dastur doirasida o'qitish imkoniyatlarini kengaytirishga yordam beradigan juda katta miqdordagi manbalar mavjud. O'qituvchi tomonidan Internet-resurslaridan olingan asl materiallardan darslarda foydalanish, haqiqiy ma'lumotlarga asoslangan turli xil ma'lumot qidirish o'qish qobiliyatlari va ko'nikmalarini shakllantirishga imkon beradi. Bundan tashqari, o'qituvchi o'quv dasturiga muvofiq maqolalarning murakkabligini aniqlaydi va o'qish bilan ishlash maqsadiga qarab mavzular va tarkibni tanlaydi. Talabalar ixtisosiga oid so'nggi yutuqlarga oid zamonaviy maqolalar, qoida tariqasida, kelajak muhandislarining qiziqishini uyg'otadi, ufqlarini kengaytiradi, darslarda keyingi ishlarini rag'batlantiradi, ilm-fan va muhandislik muammolari bilan bugungi kun aloqalarini his qilishlariga imkon beradi. Bunday materialni o'qish o'quvchilarni rag'batlantiradi, ularning kelajakdagi kasbiga tayyorgarlik holatini faollashtiradi. Talabalarning ma'lumot olish uchun mustaqil ravishda material qidirishi, uni qayta ishlash va referat tayyorlashda foydalanish til o'rganishda alohida o'rin tutadi. Shaxsiy mashg'ulotning bunday tashkil etilishi talabaga nafaqat o'zi uchun qiziq bo'lgan mavzuni tanlash, balki o'z imkoniyatlarini to'liq amalga oshirishga imkon beradigan manbaning murakkabligini aniqlash imkoniyatini ham

beradi. Shubhasiz, inshoga tayyorgarlik ko'rish uchun Internetdan foydalanish kelajakdagi muhandisni tayyorlashda turli xil vazifalarni hal qilishda yordam beradi, dunyoning so'nggi texnologiyalari bilan tanishishdan tortib, turli xil ma'lumotlarni qidirib o'qish va shu bilan yakunlanadi Veb-sahifalar va veb-saytlardan to'g'ridan-to'g'ri Internetga ulanadigan kompyuter bilan jihozlangan multimedia laboratoriyasida nafaqat o'qituvchi, balki har bir talaba uchun foydalanish til o'qitishni optimallashtirishda katta qadamdir. Bu shu qadar ko'p qirrali bo'ladiki, u yakka tartibda ishlashni, juftlik, uchlik va katta guruhlarda ishlashni, shuningdek o'quvchilarning turli darajadagi tayyorgarligini o'z ichiga olishi mumkin. Internet texnologiyalaridan foydalanish imkoniyatlari cheksizdir. Ammo, shuni unutmasligimiz kerakki, internet faqat o'qitishning yordamchi texnik vositasi bo'lib, maqbul natijalarga erishish uchun uni dars jarayonida foydalanishni to'g'ri birlashtirish zarur. Talabani bunday qimmatli va ulkan salohiyatdan foydalanishga to'g'ri va o'z vaqtida rag'batlantirish va til o'rganishda olingan resurslardan foydalanish istagini malakali ravishda yo'naltirish zarur.

Foydalanilgan adabiyotlar:

1. Ахмедов, Б. А., Шайхисламов, Н., Мадалимов, Т., Махмудов, Қ. (2021). Smart технологияси ва ундан таълимда тизимида кластерли фойдаланиш имкониятлари. Scientific progress, 1(3), 102-112.
2. Кенжабаев А.Т, Рахимов Н. Таълимда ахборот технологияси. Ўқув-услугий қўлланма. -Тошкент, 2013.
3. Talipova R. Uzluksiz ta'lim tizimida o'zbek tilini ikkinchi til sifatida o'qitish bo'yicha dasturlar. Chashma Print. -Тошкент, 2010
4. Титова, С. В. Некоторые теоретические проблемы использования компьютерных технологий в образовании / С. В. Титова // Вестник Московского университета. Сер. Лингвистика и межкультурная коммуникация. – М., 2005. – Вып. 4. – С. 39– 54.
5. <http://www.bls.gov>

CHET TILLARINI O'QITISHDA AXBOROT TEXNOLOGIYALARINING O'RNI

A.TeshaboyevADCHTI ijtimoiy-gumanitar fanlar,
pedagogika va psixologiya
kafedrası o'qituvchisi**Sayfutdinova Mexrinov Zahiddin qizi**Andijon davlat chet tillari institute
Talabasi Xorijiy til va adabiyoti
(ingliz tili) fakulteti 1-bosqich talabasi

Annotatsiya. Mazkur maqolada chet tillarini o'qitishda axborot texnologiyalarining o'rni muhim ekanligi ko'rsatilgan. Xususan, axborot texnologiyalaridan foydalanib dars mashg'ulotlarini olib borishdagi afzalliklar. Maqolada bugungi kunda yoshlarni innovatsion texnologiyalarga qiziqishi tufayli, chet tillarini ham AKT dan foydalangan holda o'rgatish haqida so'z borgan. Umuman olganda, chet tilini yangi metodlar, o'yinlar va innovatsion texnologiyalar yordamida o'rganishning samaradorligi haqida fikrlar bayon qilingan. Shuningdek, chet tilini o'rganishning bosqichlarida ham axborot texnologiyalarining o'rni ahamiyatli ekanligi yorqin ifodasini topganligini ko'rishingiz mumkin. Har bir bosqichda axborot resurslari va texnologiyalardan foydalanib mahoratni oshirish mumkinligi to'g'risida fikrlar bayon qilingan. Axborot texnologiyalardan foydalanib dars o'tishda yaxshi natijalarga erishish haqida ushbu maqolada so'z boradi. Dars jarayonlarida zamonaviy texnologiyalardan foydalanib dars mashg'ulotlarini olib boorish, albatta, samarali natija beradi. Ko'rib turganingizdek, ushbu maqolada axborot texnologiyasining ta'limdagi ya'ni o'qitish metodikasidagi o'rni haqida batafsil so'z yuritilgan.

Kalit so'zlar: Axborot texnologiyalari, innovatsion texnologiyalar, chet tili, ingliz tili, komputer, televizor, multimedia, audio, e-Book, AKT.

Annotation. This article shows the important role of information technology in teaching foreign languages. In particular, the advantages of conducting lessons using information technologies. The article talks about teaching foreign languages using ICT, due to the interest of young people today in innovative technologies. In general, opinions were expressed about the effectiveness of learning a foreign language using new methods, games and innovative technologies. You can also see that the role of information technology is clearly expressed in the stages of learning a foreign language. At each stage, ideas are expressed about the possibility of improving skills using information resources and technologies. This article talks about achieving better results in teaching using information technologies. Conducting lessons using modern technologies in the course of the lesson will definitely give an effective result. As you can see, this article discusses in detail the role of information technology in education, that is, in teaching methods.

Keywords: Information technologies, innovative technologies, foreign language, English, computer, television, multimedia, audio, e-Book, ICT.

Аннотация. В данной статье показана важная роль информационных технологий в обучении иностранным языкам. В частности, преимущества проведения уроков с использованием информационных технологий. В статье говорится об обучении иностранным языкам с использованием ИКТ, в связи с интересом современной молодежи к инновационным технологиям. В целом были высказаны мнения об эффективности изучения иностранного языка с использованием новых методик, игр и инновационных технологий. Также можно увидеть, что роль информационных технологий ярко выражена на этапах изучения иностранного языка. На каждом этапе высказываются идеи о возможности повышения квалификации с использованием информационных ресурсов и технологий. В данной статье рассказывается о достижении лучших результатов в обучении с помощью информационных технологий. Проведение уроков с использованием современных технологий. Как видите, в данной статье подробно рассматривается роль информационных технологий в образовании, то есть в методах обучения.

Ключевые слова: информационные технологии, инновационные технологии, иностранный язык, английский язык, компьютер, телевидение, мультимедиа, аудио, электронная книга, ИКТ.

Dunyo har kuni, har soniya rivojlanib bormoqda. Yildan yilga insonlar o'rtasida chet tillarini o'rganishga bo'lgan ehtiyoj ham ortib bormoqda. Tabiiyki, bu holat bugungi kunda eng ommalashgan til - ingliz tilini mavqeyi ortishiga sabab bo'ldi. Shu jumladan bizning yurtimizda ham ingliz tilini o'rganishga bo'lgan talab va e'tibor oshdi. Bu jarayon davomida axborot texnologiyalaridan foydalanish ham muhim ahamiyat kasb etdi. Maktablardagi chet tili xonalari qayta ta'mirlandi va kerakli barcha jihozlar (interaktiv doska, zamonaviy o'quv qurollar, o'yinlar uchun maxsus jihozlar, televizor, proyektor va boshqalar) bilan ta'minlandi. Bundan tashqari, milliy o'quv dasturida ham AKT dan foydalangan holda yangi o'quv darsliklar yaratilgan bo'lib, endilikda o'quv jarayonlari zamonaviy texnologiyalar bilan birgalikda olib boriladi. Bu esa ingliz tilini o'rgatishda va o'rganishda bir qancha qulayliklar yaratadi.

Axborot texnologiyalar ta'lim sohasi bilan chambarchas bog'landi. Prezidentimiz Shavkat Mirziyoyev raisligida chet tillarini o'qitish tizimini takomillashtirishga bag'ishlangan videoselektor yig'ilishida 2021-yildan chet tili o'qituvchilariga milliy va xalqaro sertifikatga ega bo'lish talabi ham kiritilishi takidlangandi. Hozirgi kunga kelib esa nafaqat chet tilini bilish sertifikati, balki zamonaviy texnologiyalardan foydalangan holda dars mashg'ulotini olib borish talabi ham qo'yildi. Agarda o'qituvchi innovatsion texnologiyalardan oqilona foydalana olsa quyidagi natijalarga erishadi:

- dars jarayonida o'quvchini e'tiborini mavzuga jalb etishi osonlashadi. Chunki hozirgi yosh avlod zamonaviy qurilmalar bilan birgalikda o'sib bormoqda. Yoshlarni innovatsion texnologiyalarga qiziqishi kuchli va bu qiziqish o'quvchini ingliz tilini o'rganish yo'liga olib kirishga sabab bo'ladi.

- sinflar kesimida innovatsion texnologiyalardan foydalanish yuqori natija beradi. Boshlang'ich sinflarda ingliz tilini multimedia, qo'shiq va grafik rasmlar orqali tashkillashtirish kichik yoshdagi maktab o'quvchisida chet tiliga bo'lgan qiziqishni orttiradi.

- 5-9- sinflar davomida chet tilidagi radioeshittirishlar, television dasturlar va kichik e-Booklar tahlil qilinadi. Ingliz tili bo'yicha umumiy tushunchalar va ingliz tili aspektlari bo'yicha

ko'nikmalar beriladi. Natijada, 10-11-sinflarda oldin o'rganilgan bilimlar asosida to'liq mustaqil shug'ullanadi, hujjatli filmlar ko'riladi va muhokama uchun muhim mavzular o'rtaga tashlanadi.

O'quvchi o'zining fikrini taqdimot ko'rinishida televizordan foydalanib ham ifodalab berishi mumkin.

Darslar o'yinli tashkil etiladi. Misol uchun, televizorda kichik ovozsiz multfilm qo'yiladi, o'quvchilar uni o'zlari mimikalar orqali sinfdoshlariga namoyish etadi. Bu holat o'quvchida talaffuzni va gapirish ko'nikmasini shakllantiradi. O'quvchini mustaqil fikrlash qobiliyatini oshirish uchun ham o'yinlar mavjud. Masalan, televizorda hikoyaning kirish qismi qo'yiladi yoki audiodan parcha eshittiriladi. So'ngra o'quvchilar hikoyani o'zlarining mustaqil fikri yordamida davom ettiradi. Bu holat o'quvchida mustaqil, to'g'ri va aniq fikrlashni ham yuzaga chiqaradi. Eng muhimi, axborot vositalaridan va innovatsion texnologiyalardan foydalanib dars o'tish, berilgan dars soati mobaynida barcha o'quvchilar bilan individual shug'ullanish imkonini ham beradi. Barcha o'quvchilar darsga faol qatnashadilar. Bir vaqtning o'zida ham o'qituvchining AKT dan foydalanish ko'nikmasi ortadi, ham o'quvchi yaxshi o'zlashtiradi.

Barcha til ixlosmandlariga ma'lumki ingliz tili 4 ta aspektlardan (o'qish, yozish, gapirish, tinglash) tashkil topgan.

O'qish - dars mobaynida bu mahoratni rivojlantirish uchun proyektordan foydalangan holda (katta ko'rinishda) maqola, badiiy asar, kichik hikoyalarni o'quvchiga ko'rsatish mumkin. Ayrim o'quvchilarda ko'rib eslab qolish qobiliyati kuchli bo'ladi va ular ko'rib o'qish orqali til o'rganadilar.

Tinglash - bu mahoratda o'quvchiga kompyuterdan foydalanib audiolar qo'yiladi. Eshitish analizatorlari orqali audioyozuv miyada analiz qilinadi. Bu holatda o'quvchi ham tinglaydi, i ham native speaker bo'lgan odamni o'z ona tilidagi so'zlarni qanday to'g'ri talaffuz qilishini ham o'rganadi. Shuningdek, "Sun'iy aql va tillarni o'rganish" maqolasida B.Skinner tilni osonlik bilan o'rganishda kompyuterning roli muhimligini ko'rsatgan. Kompyuterda nafaqat tinglash, balki yozish mahoratida ya'ni o'z ustida individual shug'ullanishda ham foydalanish mumkin.

Yozish - bu mahorat, qolganlaridan murakkab bo'lib bunda so'z boyligidan, grammatikadan to'g'ri foydalanish muhim ahamiyat kasb etadi. Ayniqsa, ijtimoiy tarmoqlarning o'rni beqiyos. O'quvchilar internet yordamida chet ellik tengdoshlari bilan ingliz tilida yozma muloqot qiladilar. Interaktiv doska orqali ham dars mobaynida yozish mahoratini oshirish imkoni mavjud. Doskalar, ya'ni interaktiv doskalar Marti tomonidan fanga kiritilgan. Dastlab grammatik mashqlarni, mantiqiy iboralarni ifodalashda ishlatilgan. Bugungi kunga kelib esa uning vazifalari oshdi.

Gapirish - bu mahoratda o'quvchining qobiliyati yuzaga chiqadi. Til ko'nikmalari qay darajada ekanligi namoyon bo'ladi. Dars jarayonida kompyuter texnologiyasi yordamida (o'yin orqali) bu mahoratni oshirish mumkin. O'quvchi kompyuterda berilgan raqamlardan istalganini tanlaydi va uning ortidagi mavzu (yoshiga va bilim darajasiga mos bo'lgan) haqida o'z fikrlarini bayon etadi. Bunday usullar o'zaro sheriklar bilan birga muhokama qilish natijasida o'quvchida sekin-asta so'zlash qobiliyatini vujudga keltiradi.

Xulosa o'rnida shuni aytish kerakki, bugungi kunda chet tillarini o'qitishda kompyuter texnologiyalari muhim ahamiyat kasb etmoqda. Informatika, raqamli texnologiyalar fanlari ham chet tilini o'qitish metodikasi bilan bog'landi. O'ylaymanki, endilikda axborot texnologiyalari orqali chet tillarini o'rganish oson va tez amalga oshadi. Buning uchun, albatta, yosh kadrlarga zamonaviy texnologiyalar yordamida dars o'tishni o'rgatish, ularni bu boradagi malakasini

oshirish lozim. Zero, chet tillarini mukammal o'rganish ilm cho'qqisiga chiqish uchun yo'l bo'lib hisoblanadi.

Foydalanilgan adabiyotlar:

1. Bekmuratova U.B. "Ingliz tilini o'qitishda innovatsion texnologiyalardan foydalanish" mavzusida referat. Toshkent - 2012 yil.
2. Djumaqulov A.A. The difference modular teaching technology and traditional teaching, 2020.
3. Jalilov J. Chet tili o'qitish metodikasi, Toshkent. 2012
4. Otaboyeva. M.P. Chet tilini o'qitishda zamonaviy innovatsion texnologiyalaridan foydalanish va uning samaradorligi / M.P. Otaboyeva. - 2017.
5. O'. Hoshimov, I. Yoqubov. "Ingliz tili o'qitish metodikasi" (o'quv qo'llanma) Toshkent: "Sharq" nashriyoti, 2003.

O'ZGA TILLI GURUHLARDA TIL O'QITISHDA MULTIMEDIA TAQDIMOTIDAN FOYDALANISH

N. Bekmuxamedova

O'zbekiston Milliy universiteti
katta o'qituvchisi

Annotatsiya: Ushbu maqola darslarda multimediya prezentatsiyalaridan foydalanish haqida, bu sizga ona tilining barcha jihatlarini: kognitiv, o'quv, rivojlanish va tarbiyaviy, ammo ta'lim jihatidan - nutq faoliyatining barcha turlarini: o'qish, gapirish, tinglash, yozishni o'zlashtirishga kommunikativ yondashuvni amalga oshirishga imkon beradi.

Kalit so'zlar: multimedia, sxema, chizma, diagramma, taqdimot, dizayn, animatsiya.

Аннотация: Данная статья об использовании на уроках мультимедийных презентаций, которые позволяют реализовать коммуникативный подход к овладению всеми аспектами не родного языка: познавательным, учебным, развивающим и воспитательным, а внутри учебного аспекта - всеми видами речевой деятельности: чтением, говорением, аудированием, письмом.

Ключевые слова: мультимедиа, схема, чизма, диаграмма, такдимот, дизайн, анимация.

Annotation: This article is about the use of multimedia presentations in the classroom, which allow you to implement a communicative approach to mastering all aspects of a non-native language: cognitive, educational, developmental and educational, and within the educational aspect - all types of speech activity: reading, speaking, listening, writing.

Keywords: multimedia, diagram, chism, diagram, takdimot, design, animation.

Materialni vizual ko'rsatishning eng keng tarqalgan usuli bu - taqdimotdir. Multimedia taqdimoti matnli materiallar, fotosuratlar, chizmalar, slayd-shoular, ovozli dizayn va hikoya, videokliplar va animatsiya, uch o'lchamli grafikalarini o'z ichiga olishi mumkin bo'lgan dasturdir. Bu ko'p funktsiyali va rasmlar, diagrammalar, videolar va boshqalarni ko'rsatish uchun mos keladi. Taqdimotga krassvordlar, musiqa yoki audio fayllarni kiritish mumkin.

Hozirgi vaqtda, u OTMdagi talabalari, maktabdagi o'quvchilari, ma'lum bir kurslarning tinglovchilari, tashkilot va muassasalarda yangi loyihani hamkasblarga, ma'lumotlarni qulay va oson idrok etish va o'zlashtirishga yordam beradi. Taqdimotlar, shunchaki, so'z yoki matnni ko'rsatish va rasm qo'shish kabi oddiy bo'lishi mumkin yoki murakkab yaratilishi mumkin.



Ushbu dastur chiziqli bo'lmagan tuzilishga ega bo'lgan yangi turdagi taqdimotlarni yaratishga imkon beradi, bu yerda barcha materiallarni bitta rasmga yig'ish yoki aksincha, har bir element kengaytirilishi va batafsilroq ko'rib chiqilishi mumkin bo'ladi. Taqdimot yaratilgandan so'ng, uni yuklab olish kerak, undan keyingi foydalanish hech qanday dasturiy ta'minot o'rnatishni talab qilmaydi.

Multimedia taqdimotlari ularning yanada innovatsion versiyasidir, chunki ular qo'shimcha ravishda tinglovchilar e'tiborini eng samarali ushlab turish uchun bir qator jihatlarni birlashtirgan dinamika, tovush, tasvirlardan foydalanishni o'z ichiga oladi. Multimedia taqdimotlari taqdim etilgan materialning aniqligi bilan ajralib turadi. Ular talabalarning ma'lum bir yoshi uchun o'qituvchi tomonidan ijodiy moslashtirilgan material deb hisoblanadi.

Bugungi kunga kelib, multimedia taqdimotlari o'quv materialini taqdim etishning eng samarali va tez-tez qo'llaniladigan usullaridan biri hisoblanadi, chunki deyarli har bir o'qituvchi ularni yaratish va o'quv jarayoniga integratsiya qilish qobiliyatiga egadir. [2].

Muzeyga sayohat



Multimedia taqdimotidan tilning barcha jabhalarini, ya'ni fonetika, lug'at va grammatikani o'rgatishda, shuningdek, o'rganishning turli bosqichlarida, masalan, yangi materialni kiritishda yoki uning o'zlashtirilishini nazorat qilishda foydalanish mumkin. Ushbu turdagi moddiy ta'minot nutq faoliyatining barcha turlarini (tinglash, gapirish, o'qish va yozish) faollashtirishga qodir. Multimediali taqdimotlardan foydalanish darsning tuzilishi va borishini buzmaydi, o'quv jarayonining barcha zarur elementlarini o'zida saqlaydi.

Taqdimot mazmunini to'g'ri tanlash, birinchi navbatda, slaydlar mazmuni o'qituvchi tomonidan aytilgan materialni takrorlamasligi kerak. Aks holda, o'quvchilarning e'tibor darajasi pasayadi, shuningdek, taqdimotning o'ziga bo'lgan qiziqishi pasayadi va shuning uchun uning darsdagi zaruriyati shubha ostiga olinadi.

Ushbu yondashuv vakillari qanday g'oyani ilgari surganlar. Olimlarning fikrlariga izoh bering.



Taqdimotning ixchamligi tanlangan materialni tizimlashtirish, yagona dizaynni ishlab chiqish va “quruq” matnni rasmlar, grafikalar, fotosuratlar, video va audio qismlar bilan suyultirishni nazarda tutadi.

Taqdimotni mantiqiy qurish uchun slaydlarni joylashtirish tartibi va taqdimotning zarur strukturaviy elementlari, ya’ni sarlavhali slayd, mundarija, o’quv materialining o’zi, xulosalar, mavzu bo’yicha axborot resurslarining mavjudligi, ayniqsa, muhimdir.

Multimedia taqdimoti PowerPoint, Prezi va Keynote kabi kompyuter dasturlari yordamida tayyorlanadi. [5]

Shunday qilib, taqdimotlarni tuzishga to’g’ri yondashish bilan ular ravshanligi, qisqaligi va ma’lumotliligi tufayli bir qator til materiallarini uzatishni sezilarli darajada osonlashtirishi mumkin, shuningdek, o’qituvchi tomonidan mustaqil ta’lim uchun tarqatma material yoki ma’lumotnoma sifatida foydalanish mumkin.

Foydalanilgan adabiyotlar:

1. Азимов Э. Г., Щукин А. Н. Новый словарь методических терминов и понятий(теория и практика обучения языкам). М.: Издательство ИКАР, 2009.
2. Бодров В.Н., Магалашвили В.В. Ориентированная на цели визуализация знаний // Educational Technology & Society. 2008. Т. 11, № 1. С. 420-433.
3. Dexkanova M.U. Ta’lim jarayonida noan’anaviy ta’lim texnologiyalarini qo’llash usullari. // Ta’lim texnologiyalari. –Toshkent: 2006. - № 1.
4. Najmiddinova M. Ta’limning sifat va samaradorligini oshirishda vizuallash-tirishning ahamiyati. // Ta’lim texnologiyalari.-Toshkent: 2006. -№1.
5. Пассов Е.И Узок иностранного языка /Пассов Е.И., Кузовлева Н.Е. - Ростов н/Д: Феникс, ГЛОССА-Пресс, 2010. 640 с.

TA'LIM JARAYONIDA KOMPYUTER TEXNOLOGIYALARI, MULTIMEDIALI O'QUV KONTENTLARIDAN FOYDALANISH O'RNI VA SAMARASI

Topildiyeva Feruza Raximjonovna
O'zbekiston Milliy universiteti o'qituvchisi

Annotatsiya: Ushbu maqolada hozirgi kun talabi bo'yicha ta'lim-tarbiya jarayonida kompyuter texnologiyalari, multimedial o'quv kontentlaridan foydalanish o'rne va samarasi masalalari yoritib berilgan. Bunda yangi pedagogik texnologiya prinsiplari izohi berilgan. Yangi pedagogik texnologiyalardan o'zbek tilini o'qitishda maqsadli foydalanish, talabalarni mavzularni o'zlashtirishda erkin va mustaqil yondashishi, ta'lim-tarbiya jarayonida qiziqarli bilim berish, ularning dunyoqarashida vatanparvarlik, o'zaro hurmat, tilga bo'lgan mehr-muhabbat, mustaqil fikrlash, xulosa qila olish ko'nikmalarni shakllantirish masalalari yoritib berilgan. Shu bilan birga maqolada zamonaviy o'qituvchi ta'limda samarali usullardan foydalanib dars o'tishi, o'zi ham yangi pedagogik va axborot kommunikatsion texnologiyalar ishlab chiquvchi va o'z mehnat faoliyatining tashkilotchisi, tanqidchisi ham bo'la olishi, u tayyorlagan mutaxassis dunyoning istalgan rivojlangan davlatidagi kadrlar bilan raqobatlasha olishi masalalari ham ochib berilgan.

Kalit so'zlar: kompyuter texnologiyalari, multimedia, o'quv kontentlari, ta'lim jarayoni, axborot -kommunikatsiya texnologiyalari.

Аннотация: В данной статье описывается роль и эффективность использования компьютерных технологий и мультимедийного образовательного контента в образовательном процессе. В статье описаны принципы новой педагогической технологии. Раскрыты такие проблемы, как: целенаправленное использование новых педагогических технологий в процессе обучения узбекского языка, свободный и самостоятельный подход студентов в освоении тем, интерактивное обучение в учебном процессе, формирование в их мировоззрении патриотизма, взаимоуважения, любви к языку, свободы мышления, умения делать выводы. В то же время в статье отмечается, что современный педагог может преподавать с использованием эффективных методов в образовании, может быть разработчиком новых педагогических и информационно-коммуникационных технологий, а может быть организатором и критиком собственной работы, что подготовленный им специалист может конкурировать с кадрами из любой развитой страны мира.

Ключевые слова: компьютерные технологии, мультимедиа, образовательный контент, образовательный процесс, информационно-коммуникационные технологии.

Annotation: This article describes the role and effectiveness of the usage of computer technology and multimedia educational content in the educational process. The article describes the principles of the new pedagogical technology. Such problems as the purposeful usage of new pedagogical technologies in the process of teaching the Uzbek language, the free and independent approach of students in mastering topics, interactive learning in the educational process, the formation of patriotism in their worldview, mutual respect, love for the language, freedom of thinking, the ability to do conclusions were discussed. At the same time, the article notes that a modern teacher can teach with effective methods in education, can be a developer of new

pedagogical and information and communication technologies, or can be an organizer and critic of his own work, a specialist trained by him can compete with specialist from any developed country of the world.

Keywords: computer technologies, multimedia, educational content, educational process, information and communication technologies.

Hammamizga ma'lum bugungi kunda, biz rivojlangan fan-texnika asrida yashayapmiz. Bu esa ta'lim- tarbiya jarayonida pedagogik texnologiyalarning to'g'ri joriy qilinishi o'qituvchining bu jarayonda asosiy tashkilotchi yoki maslahatchi sifatida faoliyat ko'rsatishiga olib keladi. Ayni damda ta'lim va tarbiya jarayonida axborot-kommunikatsiya texnologiyalaridan foydalanishning samarasi hozirgi zamon fan-texnika taraqqiyoti va axborot oqimining kun sayin o'sib borishi, ular asosida bajariladigan jarayonlarni yanada tezroq amalga oshirishni talab qiladi. Shu bois, bugungi kunda zamonaviy texnika vositalariga asoslangan kompyuter texnologiyalarini o'quv jarayoniga, jamiyat faoliyatining turli yo'nalishi va barcha jabhalariga tatbiq etish davr talabiga aylanib ulgurgan. Bugungi kunda respublikamizda ta'lim tizimining asosini yuqori sifatli texnologiyalar muhiti tashkil etib bormoqda. Bu sifatli texnologiyalar muhiti deb yuritilib, ta'lim tizimini takomillashtirishga, axborot-texnologiyalarini joriy etishga xizmat qiladi. Ayni damda, ta'lim tizimi xodimlari oldiga qo'yiladigan talab, bu boradagi dunyoqarashlarni keskin o'zgartirishni, kompyuter texnikasi va axborot-kommunikatsiya texnologiyasi vositalaridan barcha fanlarni o'qitishda, auditoriyadan tashqari o'tkaziladigan turli tadbirlarda va hayotimizning har bir sohasida foydalanishni talab etmoqda. Buning uchun, avvalo, har bir mutaxassis o'z fani bo'yicha chuqur bilimga ega bo'lishi va bu bilimlarni talabalarga yetkazib berish qobiliyatiga ega bo'lishi, o'qitishning zamonaviy vositalari kompyuter va boshqa axborot-kommunikatsiya texnologiyalari vositalaridan erkin foydalana olishi va ularni ta'lim-tarbiya jarayonida qo'llay olishini taqozo etadi. Axborot-kommunikatsiya texnologiyasidan foydalanish degani, hamma kompyuter yonida muntazam o'tirsing degani emas, aksincha, zamonaviy texnologiyalarning ta'lim tizimidagi o'rnini, ahamiyatini o'rganish, amaliyot jarayonida tatbiq etish ko'nikmasini yo'lga qo'yishdan iboratdir.

Ta'limda kompyuter texnologiyalaridan foydalanishning eng muhim jihatlari mavjud bo'lib;

- kompyuter imkoniyatlari cheksizdir, har bir talaba fanni o'zlashtirish jarayonida va kerakli bilimlarni egallashda, shaxsiy ehtiyojidan kelib chiqqan holda, qancha o'rganish talab qilinsa, shuncha o'rganishi mumkin;

- kompyuter texnologiyalari dasturlari modulli ta'lim asoslari negizida quriladi;

- talaba tomonidan alohida o'zlashtirilgan har bir o'quv fani kesimida amalga oshiriladi;

- talabalar o'quv jarayonidan so'ng, o'ziga qulay vaqtda ish jarayonidan uzilmagan holda foydalanishi mumkin;

- talabalar ta'lim muassasasidan qanchalik uzoqda joylashishidan qat'iy nazar, o'quv jarayonini tashkil qilinishiga imkon yaratadi;

- buning oqibatida tejamkorlikka erishiladi;

- talabalarining yangi vazifalari nazarda tutiladi, ularga nisbatan talablar an'anaviy ta'limdan keskin farq qiladi, mustaqil izlanishga yo'naltiradi;

- axborot-kompyuter texnologiyalarining barcha yangi dasturlaridan kompyuter tarmoqlaridan multimedia tizimi kabilardan oqilona foydalanish imkoniyatiga ega bo'ladi;

- kompyuter texnologiyalari asosida o'qitish ma'lum darajada ijtimoiy zo'riqishning oldini oladi, chunki yashash va moddiy sharoitidan qat'iy nazar hamma talabalarga bir xil bilim faoliyatini faollashtirishga sharoit yaratadi, imkon beradi;

- kompyuter texnologiyalari dasturlarini o'quv jarayoniga tatbiq etish yo'li bilan yuqorida qayd etilganlarning barchasini amalga oshirish mumkinligini ta'kidlashimiz mumkin;

- joydan unumli foydalanish imkoni yaratiladi;

- ko'rgazmalilik, bayon etish imkoni mavjud bo'ladi;

- vaqtdan unumli foydalanish imkoni, kompyuter texnologiyalari yordamida dars o'tish jarayoni davomida nazariyani amaliyotga bog'lab olib borishga sharoit yaratiladi;

- axborotni muntazam yangilab borish imkoni – bunda elektron darsliklarda yangi ma'lumotlarni bir necha daqiqalarda dasturga kirib o'zgartirish mumkin bo'ladi;

- talabalarining bilim darajalarini har tomonlama tizimli va yaxlit tekshirib ko'rish imkoni

- barcha multimedia darsliklarida talabalarining o'zlashtirganlik darajasini tekshiruvchi topshiriqlar va test savollari kompyuter yordamida avtomatik ravishda bajariladi;

- zamonaviy talabaning takomillashuv imkoni - XXI asr kompyuter texnologiyalari asri, ta'lim tizimi xodimlari, ayniqsa, o'qituvchilar davr bilan hamohang holda va turli xil kasb egalari axborot-texnologiyalarining barcha imkoniyatlaridan samarali foydalanib kelmoqdalar. O'zbek tili, o'zbek tilining sohada qo'llanishi, davlat tilida ish yuritish fanlarini o'tish jarayonida ham kompyuter texnologiyalaridan keng foydalaniladi. Bu esa kompyuterda ishlash malakasini rivojlantirish imkonini - jamiyatning har bir a'zosi, shu jumladan, talabalarni ham internet tarmoqlari va elektron manbalardan ma'lumotlarni mustaqil ravishda olishlari, ularning intellektual salohiyatini oshirishga imkon beradi, bu esa talabalarining ijodiy yondoshishiga va tanqidiy fikrlashiga, elektron darsliklar, internet tarmoqlariga kirib fanga tegishli yangi kiritilgan ma'lumotlarni izlab topishi, kuzatishi, ma'lumot yig'ishi, ularni saralashi, tartibga va tizimga solishi, tahlil qilishi nazariy bilimlarni tadqiq qilishi, natijalarni taqdimotlar yordamida asoslab berishi mumkin bo'ladi. Bunda talaba mashg'ulot jarayonining passiv tinglovchisidan faol ishtirokchiga aylanadi, yana ularda mustaqil ta'lim madaniyatini shakllantiradi. Bugungi kunda darslarni kompyuterlashtirish, elektron doskalardan foydalanish muhim ijobiy natijalarga olib kelyapti. Bulardan tashqari, multimedia texnologiyalari kompyuter yordamida bilimlarni turli ko'rinishlarda ifodalash imkoniyatiga ega, masalan:

- tasvirlar, skanerlangan rasmlar, jadvallar, chizmalar, xaritalar va slaydlar;

- ovoqli yozuv, ovoqli samaralar va musiqalar;

- video, murakkab video samaralar va animatsiya, imitatsiyalari;

- animatsiya va simulyatsiyalarni taqdim etilishi talabalarda tasavvur hosil qilib, o'tilgan mavzuni samarali o'zlashtirishga katta imkon beradi.

Ta'lim-tarbiya jarayonida axborot-kommunikatsiya texnologiyalaridan foydalanish talabalar bilim samaradorligini oshirishda muhim o'rin tutadi.

Shu bilan birga zamonaviy o'qituvchi ta'limda samarali usullardan foydalanib dars o'tishi, o'zi ham yangi pedagogik va axborot kommunikatsion texnologiyalar ishlab chiquvchi va o'z mehnat faoliyatining tashkilotchisi, tanqidchisi ham bo'la olsun, u tayyorlagan mutaxassis dunyoning istalgan rivojlangan davlatidagi kadrlar bilan ham ilmda, ham mahoratda, ham odob-axloqda erkin raqobatlasha olsun. Xullas, shunday ekan, ta'lim-tarbiya tizimiga multimediali elektron resurslarni joriy qilish pedagogikaning asosiy vazifalaridan biriga aylandi. Kompyuter asrida, fan-texnikaning barcha sohasi shiddat bilan rivojlanayotgan bir davrda, barcha sohalar

kabi o'zbek tili fanining o'rganish darajasi ham takomillashib bormoqda, bu ta'lim samaradorligini oshirishga xizmat qiladi. O'zbek tilida multimedialar, elektron darsliklarning yaratilishi tillarni o'qitish jarayonida multimediali o'quv kontentlarini kengayishiga olib keladi. "Shu kunga qadar ham o'zbek tilshunoslik ilmining zamonaviy yo'nalishlariga o'z hissasini qo'shgan A. Jo'rayev, o'zbek sotsiolingvistikasi asoschisi, A.Pulatov kompyuter lingvistikasiga munosib hissa qo'shgan olim sifatida, tilimizni nufuzini rivojlantirishda yana bir qator olimlar; R.Sayfullayeva, B.Mengliyev, M.Qurbonova, D.Lutfullayeva, M.Yuldashevlarni e'tirof etmoq o'rindir." [1] Zamonaviy texnologiyalarga asoslangan darsliklarni yaratish uchun o'zbek tilida milliy qadriyatlar aks etgan vatanparvarlik, oilaparvarlik, milliylik, mehnatsevarlik, ona tilimizga bo'lgan hurmat hissi yo'g'rilgan g'oyalar singdirilgan, matnlar videoroliklar, sohaga doir materiallar tayyorlanib mavzular bo'yicha tizimga solinyapti. Albatta, innovatsion texnologiyalar asosida yaratilyapti. Bu esa dars mashg'ulotlarida, har bir mavzuni innovatsion texnologiyalar asosida qiziqarli tarzda o'tishga keng yo'l ochadi.

AKTlar asosida yangi loyihalarni yaratish hamda boyitish imkonini beradi. Bugungi kunda ta'lim jarayonida kompyuter texnologiyalari va multimediali o'quv kontentlaridan foydalanish o'zni va samarasi sezilarli darajada rivojlanib bormoqda.

Foydalanilgan adabiyotlar:

1. Abdurahmonova.M. O'zbek tilshunosligidagi zamonaviy yo'nalishlar. Ilmiy-amaliy anjuman materiallari. Toshkent. 2020.
2. Ботвенко М.А. Компьютерная лингводидактика. / Учебное пособие.М. 2009.
3. Saidahmedov N. Yangi pedagogik texnologiyalar – T.: "Moliya". 2003.
4. Полат Е.С. Новые педагогические и информационные технологии в системе образования., - М.: 1999.
5. Usmanov A.I. Zamonaviy axborot texnologiyalari. T.: № "Akademiya", 2007.

TILLARNI O'QITISH JARAYONIDA MULTIMEDIALI O'QUV KONTENTLARIDAN FOYDALANISH

Fattaxova Diloram Abduraxmanovna

p.f.n. O'zbekiston Milliy universiteti

Annotatsiya: Maqolada universitetda til fanlarini o'qitishda zamonaviy axborot-kommunikatsiya texnologiyalaridan foydalanish masalalari ko'rib chiqiladi. Ta'limning o'ziga xos xususiyatlariga moslashtirilgan zamonaviy texnologiyalar, multimedia materiallari, audio va video darslardan, taqdimotlardan foydalanish o'zbek tilining amaliy kursini o'rganishda talabalarni tayyorlash darajasi va sifatini oshirishga imkon beradi.

Kalit so'zlar: multimedia texnologiyalari, axborot-kommunikatsiya texnologiyalari, animatsiya, pedagogik texnologiyalar, slaydlar, figuralar.

Аннотация: В статье рассматриваются вопросы использования современных информационно-коммуникационных технологий при обучении языковым дисциплинам в вузе. Использование современных технологий, мультимедийных материалов, аудио и видео уроков, презентаций, адаптированных к специфике обучения позволяют повысить уровень и качество подготовки студентов при изучении практического курса узбекского языка.

Ключевые слова: мультимедийные технологии, информационно-коммуникационные технологии, анимация, педагогические технологии, слайды, фигуры, тропы.

Annotation: The article deals with the use of modern information and communication technologies in teaching language disciplines at the university. The use of modern technologies, multimedia materials, audio and video lessons, presentations adapted to the specifics of education can improve the level and quality of students' preparation when studying the practical course of the Uzbek language.

Keywords: multimedia technologies, information and communication technologies, animation, pedagogical technologies, slides, figures, trails

Ta'lim-tarbiya jarayonida pedagogik texnologiyalarning to'g'ri joriy qilinishi o'qituvchining bu jarayonda asosiy tashkilotchi sifatida faoliyat ko'rsatishiga olib keladi. Bu esa talabadan ko'proq mustaqillikni, ijodkorlikni va irodaviy sifatlarini talab etadi. Ma'lumki, o'quv jarayonidagi pedagogik texnologiya – bu aniq pedagogik ketma-ketlikdagi yaxlit jarayon bo'lib, u talabaning ehtiyojidan kelib chiqqan holda bir maqsadga yo'naltirilgan, oldindan puxta loyihalashtirilgan va kafolatlangan natija berishga qaratilgan pedagogik jarayondir. Shu bois, mamlakatimizda kompyuter yordamida til o'rgatish ishlari tobora jadallashtirilmoqda. Turli manbalardagi ma'lumotlar va tajribalarga tayangan holda shuni aytish mumkinki, ta'lim jarayonini axborotlashtirish, xususan, oliy maktabning ta'limi rus tilida olib boriladigan guruhlarida o'zbek tili darslarini kompyuterlashtirish muhim ijobiy natijalarga olib keladi, jumladan:

- ta'lim-tarbiya jarayoni faollashtirilishi, jadallashtirilishi tufayli nutq o'stirish samaradorligi oshmoqda;
- axborotlarning turli shakllarda (ovoz, matn, video, grafika, animatsiya yordamida) berilishi til o'rganuvchilarda katta qiziqish uyg'otib, mavzuning xotirada uzoq vaqt saqlanishini ta'minlaydi;
- talaba mashg'ulot jarayonining passiv tinglovchisidan faol ishtirokchiga aylanadi;
- mustaqil ta'lim madaniyati shakllantiriladi;
- kompyuterning o'zi ham o'rganish obyekti bo'lgani bois, kompyuter savodxonligini chuqurroq o'zlashtirishga yordam beradi;
- vaqt tanqisligi muammosi sezilarli darajada kamayadi;
- o'qituvchining ko'pgina tashkiliy vazifalari kompyuter zimmasiga o'tadi;
- kompyuter talabani bilimini, mavzuni o'zlashtirish darajasini obyektiv tarzda nazorat qilib baholaydi.

Ayniqsa, talabani o'z natijalarini shu zahoti monitorda ko'rishi muhim ahamiyatga ega. Bunda rag'batlantiruvchi yoki ogohlantiruvchi elementlar talabaga katta ruhiy ta'sir o'tkazadi, xatoni tuzatishga, diqqat e'tiborini jamlashga ko'maklashadi. Mashg'ulot mobaynida bajarilishi mumkin bo'lgan mashq, vazifalar soni keskin ko'payib, mashqlar didaktik tamoyillarga mos keladi. Kompyuter sun'iy ravishda "til muhiti"ni yaratadi. Bundan tashqari, multimedia texnologiyalari kompyuter yordamida bilimlarni turli ko'rinishlarda ifodalash imkoniyatiga ega, jumladan:

- tasvirlar, skanerlangan rasmlar, chizmalar, xaritalar va slaydlar;
- ovozli yozuv, ovozli samaralar va musiqalar;
- video, murakkab videosamaralar va animatsiya imitatsiyalari;
- animatsiya va simulyatsiyalarni taqdim etilishi talabalarda tasavvur hosil qilib, o'tilgan mavzuni samarali o'zlashtirishga katta yordam beradi.

Aynan shuning uchun ham axborot va telekommunikatsion texnologiyalarning rivojlanishi bilan shaxsga yo'naltirilgan o'qitish uchun mo'ljallangan muhim pedagogik vositalar – internet, multimedia kabi ta'lim ashyolari va gipermatnlik elektron darsliklardan iboratdir.

Telekommunikatsiya va axborot texnologiyalarini ta'limda qo'llanilishi yanada rivojlangan axborot vositalarini – gipermedia tizimlarini yaratishga olib keladi. Demak, multimedia foydalanuvchiga yanada samaraliroq ta'sir etishga qaratilgan turli dasturiy va texnik vositalarni axborot texnologiyalarining spektrini ifodalaydi. O'qitishning bunday usuli texnologik yondashuv – reproduktiv ta'limga xos bo'lib, quyidagilarni nazarda tutadi, ya'ni:

- o'qish-o'qitish jarayonini o'zaro uzviy bog'liq bosqichlar, fazalar, amallarga ajratish;
- ta'limdan kutilgan natijaga erishish uchun bajariladigan harakatlarni muvofiqlashtirish, ketma-ketlikda, bosqichma-bosqich amalga oshirish;
- loyihalashtirilgan ishlar, amallarning barchasini birdek bajarish reproduktiv daraja uchun pedagogik texnologiya usulidagi ta'lim undan kutiladigan natija ham belgilab qo'yiladi, o'quv materialini aniq ifodalangan o'quv maqsadiga mos qayta tuziladi, ayrim modullarga ajratiladi, o'quv materialini o'rgatishning alternativ usullari nazarda tutiladi. O'qitishni etalonda ko'rsatilgan natijaga yetkazish maqsadga muvofiq bo'ladi. Bu holatda o'qitishga ijodiy yondashuv talab qilinadi.

Ta'lim muassasalarining ta'limiy-tarbiyaviy faoliyatida ijodiy individual yondashuvni tatbiq qilish ta'lim-tarbiyaning rivojlanuvchi modelini o'z ichiga oladi. Modomiki shunday ekan, ta'lim-tarbiya tizimiga multimedial elektron resurslarni joriy qilish pedagogikaning asosiy vazifalaridan biriga aylanadi. Kompyuter asri deb ataladigan, fan va texnikaning barcha sohalar shiddatli ravishda rivojlanayotgan bir davrda barcha sohalar kabi o'zbek tili fanini o'rganish darajasi mezonlari ham rivojlanib, takomillashib bormoqda. Shunga ko'ra, o'zbek tili fanidan yaratilayotgan multimedial o'quv resurslari talablarning og'zaki va yozma nutq, malaka hamda ko'nikmalarini shakllantirish, mazmunini boyitishga qaratilgan bo'lib, ta'lim samaradorligini oshirishga katta hissa qo'shadi.

Zamonaviy texnologiyalarga asoslangan multimedia darslarini yaratish uchun o'zbek tilida milliy qadriyatlar aks etgan, vatanparvarlik, milliylik g'oyalari singdirilgan, matnlar, videoroliklar, ixtisoslikka oid materiallar tayyorlanib, mavzular bo'yicha tizimga solinadi. Barcha materiallar mavzular bo'yicha mantiqiy izchillikda ishlab chiqiladi. Bunda asosiy e'tibor nutqning kommunikativ ko'nikmalariga, mavzularni innovatsion texnologiyalar asosida o'tishga qaratilgan bo'lib, zamonaviy AKTlar asosida media darslar yaratish uchun asos vazifasini bajaradi.

Multimedia texnologiyasi multimedial dasturiy mahsulotning interfeysi va imkoniyatlaridan foydalanish hamda aniqlash jarayonidir.

Multimedia (ingl: multimedia, multi – ko'p, media - muhit) – bir vaqtning o'zida turli shakllarda uzatiladigan ma'lumotlar yig'indisi. [1] Multimedia katta hajmdagi ma'lumotlarni saqlash va ularga tezkor kirishni ta'minlaydigan texnologiya.

Multimedia texnologiyasi quyidagilarni o'z ichiga oladi:

Kompyuterli animatsiya;

Multimedial taqdimot va interfeys yaratish;

Ovoz va video bilan ishlash.

Multimedia texnik vositalari – multimedial ma'lumotlarni qayta ishlashda qo'llaniladigan maxsus apparat vositalari bo'lib, ularga quyidagilar kiradi:

Ovoz yozish qurilmasi

Ovoz chiqarish qurilmasi

Manipulyatorlar

Virtual borliq qurilmalari

Uzatish qurilmalari

Yozib olish qurilmalari

Tasvirlarni qayta ishlash qurilmalari.

Multimedia mahsulotlari turlari quyidagilar:

Taqdimotlar

Antimatsion roliklar

O'yinlar

Video mahsulotlar

Multimedia galereyalari

Audio mahsulotlar

Web ilovalar.

Multimedianing hozirgi kunda keng va samarali qo'llanadigan sohalar:

Ta'lim

Tibbiyot
Harbiy soha
Biznes va menedjment
San'at va ijod
Arxiv ishlari
Sun'iy intellekt tizimlari
Virtual borliq va boshqalar.

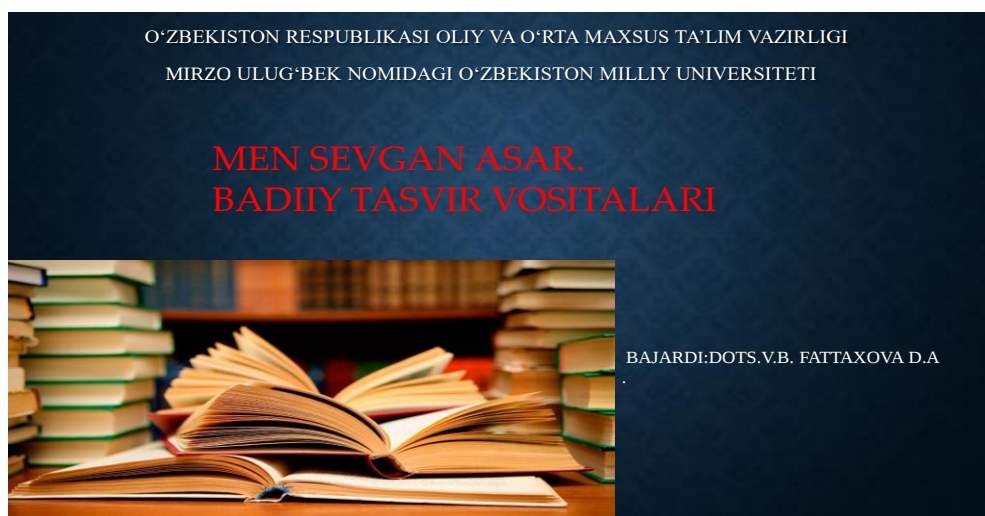
Ta'limda 1986 yil birinchi multimedialli entsiklopediya yaratildi. (Grolier). O'qish samaradorligi 30% foizga oshdi.

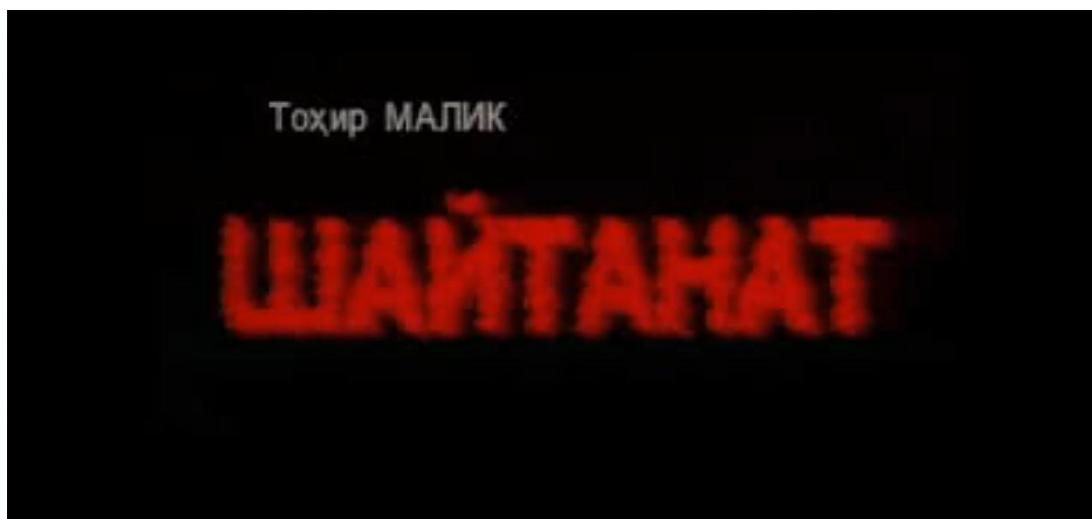
Interaktiv test tizimi.
Ta'limda multimedialli mahsulotlar turlari.
Elektron darsliklar
Elektron entsiklopediyalar
Bilimni onlayn testlash va baholash tizimi
Trenajorlar (simulyatsiya modellari)
Chet tillarini o'qitish tizimlari.

Ilmiy tadqiqot ishimizning bo'limi sifatida pedagogik texnologiyalarga tayangan holda o'zbek tilini o'rgatishni multimedia texnologiyalaridan foydalanib, o'quv kontentlari yaratishni maqsad qilib oldik.

Pedagog dars faoliyati davomida mavzularni o'tilgandan keyin, mavzuni mustahkamlash maqsadida quyida keltirilgan multimedialardan foydalansa maqsadga muvofiq bo'ladi. Multimedialda savollar, matn, audio eshittirish, mavzuga oid turli suratlar, mavzuni mustahkamlash maqsadida testlar, turli metodlar bilan boyitilgan[4]

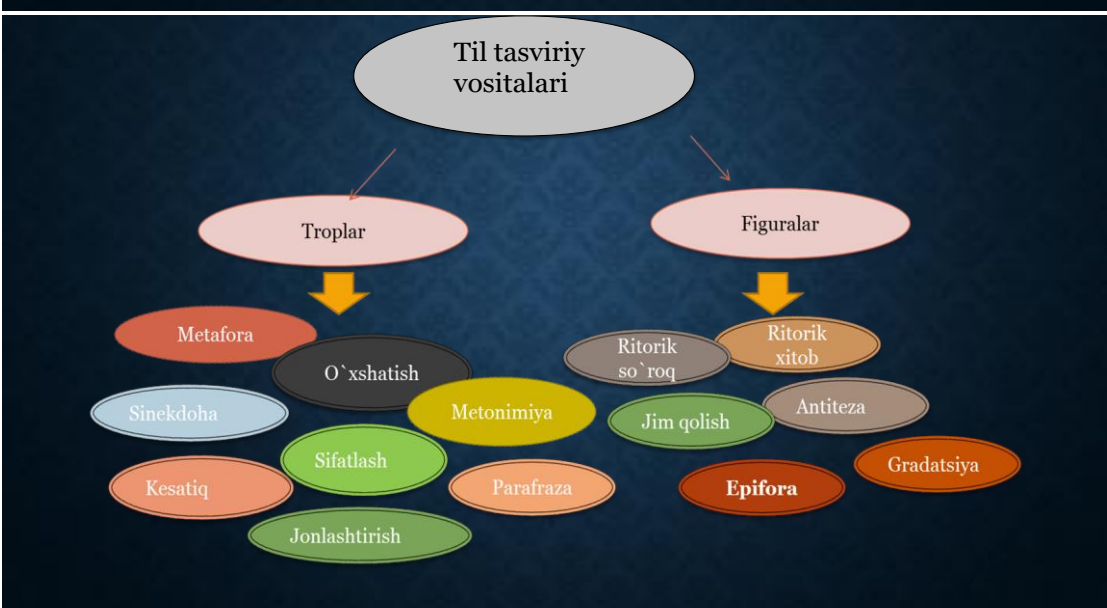
Leksik mavzular bilan tanishganda aniq va samarali vosita bo'lishi uchun multimedia juda yaxshi kompyuterlashtirilgan o'qitish texnologiyasi turi hisoblanadi. Shuning uchun ham ilmiy tadqiqot ishida pedagogik texnologiyalarga asoslangan tartibda multimedialli o'quv kontentlari yaratishni asosiy maqsad qilib belgilab oldik. Masalan, Men sevgan asar. Badiiy tasvir vositalari mavzusini ochib berishda, biz audiodars tayyorladik va talabalarga taqdim etdik. Tayyorlagan taqdimotdan namunalar keltiramiz:





Trop (gr.aylanish soʻzidan) - koʻchim, trop. Biror narsa yoki hodisani ifodalash uchun soʻz yoki soʻz birikmalarining oʻz maʼnosidan boshqa maʼnoda qoʻllanishini anglatuvchi tushuncha. Koʻchimlarning badiiy nutqda ishlatilishi ifodalilikni, taʼsirchanlikni oshirishga, tasvirlanayotgan voqea-hodisani ravshanroq baholashga yordam beradi. Koʻchimlar jonli til bilan birga, badiiy adabiyotda ham gʻoyat keng qoʻllanadigan maxsus tasviriy vositalardir.

Troplarga:epitet, oʻxshatish, metafora, metonimiya, sinekdoxa, mubolagʻa va kichraytirish, allegoriya, ironiya(kesatiq), perifraza, jonlantirish va h.k.kiradi.



Metafora (grek.- koʻchirish) koʻchma maʼno hosil qilishning eng keng tarqalgan usullaridan biri boʻlib, mumtoz adabiyotshunosligimizda “istiora” deb yuritilgan. Narsa-buyum, voqea va hodisalar oʻrtasidagi oʻzaro oʻxshashlikka asoslangan, nutqiy taʼsirchanlikni oshiruvchi vositalardan biridir.

Metafora misollar:

Agar hayot boʻston esa bogʻbon muallim
 Agar hayot karvon esa sarbon muallim
 Agar hayot uzuk boʻlsa koʻzi muallim
 Agar hayot shoir boʻlsa soʻzi muallim....(Shuhrat)

Metonimiya- yunoncha soʻz boʻlib, “yangi nom qoʻyish”, “nomini oʻzgartirish”, “qayta nomlash” maʼnolarini ifodalaydi. Ikki narsa oʻrtasidagi aloqadorlik asosida metonimiya hosil boʻladi.

Metonimiya

misollar:

Muhammad Yusufni qoʻlga olganimda, tong yorisha boshlagan edi.

Kech boʻlishiga qaramay, samovarda odam koʻp.

Jonlantirish – insonga xos xususiyatlarni jonsiz va mavhum narsalarga, hodisalarga nisbat berib tasvirlashdir.

Jonlantirish

misollar:

Soʻylasin Afrosiyobu

Soʻylasin Oʻrxun xati

Koʻhna tarix shodasida

Bitta marjon Oʻzbekim. (E.Vohidov)

Oq yoʻrgakka oʻragansan oʻzing bizni,

Ham oq yuvib – taragansan oʻzing bizni.

Beshingimiz uzra bedor ona boʻlib,

Kunimizga yaragansan oʻzing bizni.

Fidoying boʻlgaymiz seni, Oʻzbekiston,

Hech kimga bermaymiz seni, Oʻzbekiston!..

(M.Yusuf)

O‘xshatishlar badiiy nutqda keng qo‘llaniladi. Nutqning ta’sirchanligini oshirishda muhim omil sanaladi. Nutqimizda –dek (-day) qo‘shimchasi va o‘xshash, xuddi, go‘yo, misli, misoli, singari, yanglig‘, bamisoli kabi so‘zlar yordamida o‘xshatishlar yuzaga keladi.

O‘xshatish

misollar:

**Chiroylidir go‘yo yosh kelin,
Ikki daryo yuvar kokilin...**

Tog‘lardagi qip-qizil lola
Bo‘lib go‘yo yoqut piyola...(H. Olimjon)

Tabiat saharda misli bir g‘uncha,
Tong esa g‘unchaning ochilgan chog‘i.
Yoki taxmin qiling: tabiat bir qiz,
Tong esa u qizning marmar yonog‘i. (Uyg‘un).

Figura so‘zi “qiyofa, tashqi ko‘inish, obraz” ma’nolarini anglatadi.

Figuralar (stilistik figuralar, stilistik usullar, ritorik figuralar) tinglovchiga umumiy hissiy ta’sir ko‘rsatuvchi, nutqning mana shu ta’sirini oshiruvchi maxsus sintaktik qurilmalardir. Bularga: antiteza, gradatsiya, ritorik so‘roq, ritorik xitob, ko‘p bog‘lovchilik va bog‘lovchisizlik, jim qolish, takror va h. kiradi.

Antiteza – (gr. qarama-qarshi qo‘yish ma’nosini anglatadi) mumtoz adabiyotda tazod deyiladi. Antitezada, odatda, ifodalilikni kuchaytirish, qiyosni bo‘rttirish uchun qarama – qarshi tushunchalar, fikrlar, shakllar, belgilar, xarakter, xususiyatlar, miqdor yoki boshqa jihatlar ataylab zidlashtiriladi. O‘zbek tilida barcha antonim so‘zlar tazod san’atini shakllantirishda ishtirok etishi mumkin: yaxshi – yomon, oq –qora, katta – kichik, chiroyli – xunuk.

O‘zbek xalq maqollarida tazoddan keng foydalanilganligi kuzatiladi:

Torga – tor dunyo, kengga – keng dunyo.
Kattalar hurmatda, kichiklar izzatda.

Anafora - yuqoriga ko‘tarilish degan ma’noni bildiradi. Bunda har bir misra, abzats va gap boshida bir xil takrorlanuvchi konstruksiyalar ishlatiladi. Bular so‘z, so‘z birikmasi, gap shaklida bo‘lishi mumkin.

Anafora misollar:

U alla aytmasa, tonglar otmaydi,
U alla aytmasa, quyosh botmaydi... (S.Zununova)

....Ona mehri bilan to‘liqdir,
Ona mehri bilan ulug‘dir. (H. Olimjon)

Taqdimotda asardan parcha keltirilib, muallifning ijodi haqida qisqacha ma’lumot beriladi. Yangi mavzu misollar yordamida ochib beriladi. Mavzuni tushuntirishda zamonaviy pedagogik texnologiyalardan unumli foydalaniladi. Shundan so‘ng talabalar berilgan topshiriq va mashqlarni bajaradilar.

AKTning bu kabi turi darslarda faollik va optimallikni rivojlantirishda samarali metod hisoblanadi.

Bugungi kunda informatsion texnologiyalar, elektron texnologiyalar, multimedia texnologiyalarni yaratish mutaxassislik fanlariga asoslangan tartibda, mustaqil tarzda o‘zbek tilini o‘rgatishda kompyuterlashtirilgan o‘qitish texnologiyalari qudratiga ega.

Foydalanilgan adabiyotlar:

1. Деникин А. А., [Мультимедиа и искусство: от мифов к реалиям](#) // Художественная культура. Москва - 2013. — Выпуск № 1 6-стр
2. <https://prezi.com/brdtg4hgag55/presentation/>... Таълимда мультимедиа тизимлари ва масофавий ўқитиш методлари. by on 8 September 2015.
3. Полат Е.С. Новые педагогические и информационные технологии в системе образования., - М.: 1999. 224с.
4. Саидахмедов Н. Янги педагогик технологиялар - Т.: “Молия” 2003. 172-бет.
5. Ahmedova M. O‘zbek tili darslarida yangi texnologiyalardan foydalanish usullari. – Toshkent: Universitet, 2006.
6. Граудина Л. К. Грамматическая правильность русской речи. – М.: АСТ, 2004. – С.55.
7. Выразительные средства современной русской речи. Тропы и фигуры. Терминологический словарь. – М.: Феникс, 2007. –С.869.
8. Бегматов Э. Нутк маданияти тушунчаси ҳақида // Ўзбек тили ва адабиёти. –Тошкент, 1975. №5. – Б.50-58.

АКАДЕМИК ЁЗУВ ДАРСЛАРИДА ЗАМОНАВИЙ СМАРТФОНЛАРДАН САМАРАЛИ ФОЙДАЛАНИШ

Данияров Бахтиёр Худайбердиевич

Муҳаммад ал-Хоразмий номидаги ТАТУ Нурафшон
филиали гуманитар фанлар кафедраси профессори,
филология фанлари доктори

Annotatsiya: Ushbu maqolada biz ta'lim jarayonida smartfonlardan foydalanishning maqsadga muvofiqligi, nofilologiya fakulteti talabalari uchun Yevropa va Rossiya ta'lim tizimida mavjud bo'lgan, o'zbek ta'limiga joriy etilishi lozim bo'lgan "Akademik yozuv" fani haqida so'z yuritamiz.

Kalit so'zlar: axborot texnologiyalar, smartfon, Akademik yozuv, sinonimlar, noutbuk.

Аннотация: В данной статье речь идёт об целесообразности использования смартфонов в обучении, для студентов нефилологических факультетов, о предмете «Академическое письмо», существующее в Европейском и Российском обучении, которое надо внедрить в узбекское обучение.

Ключевые слова: информационные технологии, смартфон, академическое письмо, синонимы, ноутбук.

Annotation: In this article, we are talking about the advisability of using smartphones in education, for students of non-philological faculties, about the subject "Academic writing", which exists in European and Russian education, which should be introduced into Uzbek education.

Keywords: information technology, smartphone, academic writing, synonyms, laptop.

XXI asrda axborot texnologiyalari mamlakatimizni rivojlantirishda muhim o'ringa ega bo'ldi. Masalan, bundan 20-25 yil oldin sensorli smartfon telefonlarni tashvish ham qilmaganmiz. Ta'limda axborot texnologiyalari, xususan, ayni texnologiyalari asosida yaratilgan zamonaviy smartfonlar ham, birinchi navbatda ta'lim sifatini va samaradorligini oshirish bilan birga joriy qilingan fanning jozibadorligini hamda vaqtidan unumli va samarali foydalaniшни ham ta'minlash lozim.

Masalan, kancaliyari jihozlaridan (A4 formatli ko'guz, daftar, chizgich, elim va hozir) foydalanish holida o'quv rejasida mavzularga mos ravishda uslubiy tarkatma materiallarni tayyorlash davri allaqachon o'tib borgan bo'lsa-da, hanauzga qadar ayrim ta'lim muassasalarida professor-o'qituvchilardan tarkatma materiallarni shunday eski uslubda tayyorlab tashkil qilishni talab qilishadi. Vaholanki, eski uslubda o'quv-metodik tarkatma materiallarni tayyorlash o'qituvchilar vaqtining samarasiz o'tishiga olib kelib, pedagoglarda norozilik kalfiyatini ham kelib chiqishiga sabab bo'lishi mumkin.

Bunday holatlar, ya'ni ta'lim muassasalarida barcha darslarining, shuningdek akademik o'quv fanidan mashg'ulotlarining sermazmun va qiziqarli tashkil qilinishida "aqli" telefonlarining o'rnini bekiys, xususan, smartfonlar vositasida internetdan dars jarayoni uchun zarur bo'lgan har qanday mavzular bo'yicha materiallarni, shu jumladan, rang-

баранг ва турли хил кўринишдаги электрон таркатма материалларни тайёр ҳолда бир зумда ташкил қилиш мумкин.

Академик ёзув фани Ўзбекистоннинг таълим муассалари учун янги фан бўлиб, Тошкент ахборот технологиялари университети ва унинг филиалларида яқин йиллардан бошлаб мажбурий фан сифатида талабалар учун ўтилмоқда. Бу фан Америка қўшма штатлари ва Ғарбий Европа давлатлари тажрибасидан келиб чиққан ҳолда Россиянинг бир қатор олий таълим муассаларининг ўқув режаларига рус тилида Академическое письмо (“Академик ёзув”) номи билан мажбурий фан сифатида 2000 йилларнинг бошларида киритилди [4].

Рус тилида Академик ёзув фанига оид алоҳида китоб кўринишига эга илмий ёки ўқув-методик адабиётлар кам, ўзбек тилида - деярли йўқ. Борлари ҳам барча соҳа ва мутахассислари учун тилшунос олимлар томонидан умумий тарзда ёзилган, чунки ушбу фанни олий маълумотли филология йўналишидаги зурур билим, кўникма ва малакаларга эга мутахассислар ўқитиши керакли тегишли ҳужжатлар билан белгилаб қўйилган. Бу эса нофилологик йўналишдаги соҳа ва мутахассисликларни эгаллашни азм қилган талабалар, масалан, Тошкент ахборот технологиялари университети талабалари учун академик ёзув фанини ўрганиш жараёнида катта муаммо ва қийинчиликлар туғдиради.

Масалан, “Матн ва унинг турлари” (ўқув режадаги 3-мавзу) мавжуд адабиётлар ва ўқув-методик мажмуаларда қуйидагича ёритилган:

“Матнлар тарихан шаклланган адабий тилнинг турли кўринишлари бўлиб, улар фикрни ифода этиш жараёнида нутқ тузилишининг ўзига хослиги билан фарқланади”. Матнлар тараққиётини адабий жанрлар тараққиётидан ажратган ҳолда қарамаслик керак. Чунки нутқий воситалар таркиби ва уларни қўллаш меъёрлари асар жанри билан алоқадордир.” [3].

Тасаввур қилинг-а, дастурий инжиниринг соҳасини эгалламоқчи бўлган талаба матн ҳақидаги бу билимларни эшитиб, ёки ўқиб қай аҳволга тушии мумкин. Зеро матн ҳақидаги бундай билимлар, бундай ахборот технологияларини эгалламоқчи бўлган талабалар учун эмас, балким тилшунослик соҳасининг етук билимдони бўлишни қарор қилган ёшлар учун мўлжалланган.

Айрим филологларда эътироз пайдо бўлиши мумкин. Она тили ва адабиёти домлалари ахборот технологияларига оид фанлардан махсус билимларга эга эмаслар-ку!

Ана шу ерда академик ёзув фанини ўқитаётган домлалар учун ахборот технологиялари, интернет, шу жумладан, смартфонлар асқотади. Ушбу фанни ўтаётган домла матнларни, ўқув-методик мажмуаларини талабаларнинг эгалламоқчи бўлган соҳа йўналишлари бўйича тайёрлашлари, топшириқларни ҳам шунга монанд талабалар учун тушунарли, содда ва раво, илмий асосланган, лекин оммабоп услубда етказишлари зарур.

Масалан, академик ёзув фани домласи талабаларга синонимлар тушунчасини етказмоқчи бўлса қуйидагича:

“Тадқиқот натижалари асосида ўзбек тили лексик синонимларини асосий денотатив маъноси бўйича мос, лекин коннотатив маъно компонентлари хусусиятлари ва қўлланиши(функционал имкониятлари)нинг ўзига хослиги билан фарқланадиган синонимлар деб тавсифлаш мақсадга мувофиқ. Синонимлар лексик ва контекст бирикуви ҳамда деривациявий хусусиятлари ва имкониятлари билан ҳам фарқланади” [1] – тарзида эмас, чунки синонимларнинг бундай таърифи ахборот технологиялари соҳасини танлаган

талабалар учун мураккаб бўлиб, улар учун мўлжалланмаган. Лексик синонимларнинг бу илмий таърифи ўзбек тилшунослиги соҳасида қалам тебратган олимлар, соҳа мутахассислари учун тушунарлидир.

Компьютер муҳандиси бўлишни ният қилган илму толибларга домлалари синонимлар таърифини содда ва тушунарли, улар учун жозибали тарзда: “синонимлар шакли ҳар хил, маъноси эса бир-хил сўзлар”, ёки “синонимлар маъноси жуда яқин бўлган сўзлар”, ёки “синонимлар – бу маънодош сўзлар”, деб оддий ва лўнда қилиб айта қолишлари мақсадга мувофиқ [2].

Бугунги кунда смартфонга эга бўлмаган талабалар йўқлигини инобатга олган ҳолда мустақил иш топшириқлари ҳам талабаларда катта қизиқиш ўйғотадиган тарзда берилса, академик ёзув фанининг жозибадорлиги талабалар наздида анча ошади, уларнинг ушбу фанга (ўз соҳасига оид фанларга ҳам) бўлган қизиқишлари ортади, билимлари мустаҳкамланади. Чунки ахборот технологияларни пухта эгаллашни мақсад қилган ёшлар қўлда ручка билан ёзиш-чизишни, мулойимроқ қилиб айтганда унчалик хушламайди [5].

Масалан, фаол талабалар томонидан “Академик ёзув ва ахборот”, “Ахборот технологиялари соҳасида ахборот тушунчаси”, “Мен орзу қилган ноутбук”, “Замонавий компьютер қурилмалари”, “Смартфонларнинг сўнгги авлодлари”, “Интернет буюмлар нима?” “Блокчейн инновацион маълумотлар базасида интернет буюмларни сақлаш?” мавзуларида смартфонлар ёки ноутбуклар асосида катта иштиёқ билан тайёрланган *конспект* (ўқув режадаги 7-мавзу), *эссе* (12-, 13-мавзулар), *реферат* (14-, 15-мавзулар), *курс ишлари* (17-мавзу), тақдимотлар (26-, 27- мавзулар) толиби илмларда зўр қизиқиш ўйғотиб, ушбу фандан кучли таассурот билан олган билимлари хотираларида мустаҳкам ва бир умрга ўрнашиб қолади.

Foydalanilgan adabiyotlar:

1. Данияров Б. Х. Ўзбек тили лексик синонимларининг лисоний ва нутқий муносабати, лексикографик тадқиқи. Монография.- Тошкент: “Сано стандарт”, 2019. – 200 б.
2. Данияров Б.Х. Комплексный анализ семантической структуры и функционирования лексических синонимов. Монография / Б. Х. Данияров. – Ташкент: Издательство Bookmany print, 2022. – 156 с. – С. 120-121.
3. Ж.Лапасов. Бадиий матн ва лисоний таҳлил.- Тошкент: “Ўқитувчи”, 1995. - Б. 5.
4. Смирнова Н. В. К вопросу об академической грамотности // Иностранные языки: материалы конференции. СПб.: Изд-во РГПУ им. А. И. Герцена, 2013. С. 55.
5. Данияров Б.Х. Современная методика комплексного анализа семантической структуры и функционирования лексических единиц. Традиции и новации в профессиональной подготовке и деятельности педагога: сб.науч.тр. Междунар. Науч. –практ.конф. 25-27 марта 2021 г. Тверь –С.187.

ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННО КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ ПРИ ОБУЧЕНИИ ЯЗЫКУ

М.Х. Ахмедова

к.п.н. доцент, Национальный университет Узбекистана

Фаттахова Дилорам Абдурахмановна

к.п.н., Национальный университет Узбекистана

Annotatsiya: Maqolada universitetda til fanlarini o'qitishda zamonaviy axborot-kommunikatsiya texnologiyalaridan foydalanish masalalari ko'rib chiqiladi. Sifatli ta'limning asosiy ko'rsatkichlaridan biriga aylangan axborot kompetensiyasi zamonaviy ta'limning asosiy shartlaridan biri bo'lib, uning shakllanishi bevosita o'quvchilarning ochiq axborot-ta'lim muhitidagi faol faolligiga bog'liq. Zamonaviy axborot texnologiyalari til fanlarini o'qitishda individual ta'lim traektoriyalarini ishlab chiqish va tartibga solish uchun noyob didaktik sharoitlarni yaratishga imkon beradi. Zamonaviy o'qitish texnologiyalaridan talabalarining o'rganish xususiyatlariga moslashtirilgan multimedia materiallari, shuningdek bepul mavjud bo'lgan o'quv resurslari bilan birgalikda foydalanish rus tilining amaliy kursini o'rganishda talabalarining motivatsiyasi, darajasi va tayyorgarligi sifatini oshirishi mumkin. Motivatsiyani shakllantirishda talaba tomonidan olingan ma'lumotlarni taqdim etishning yangiligi va shakli, shuningdek ularni izlanish, ijodiy faoliyatga kiritish alohida rol o'ynaydi.

Kalit so'zlar: axborot-kommunikatsiya texnologiyalari, til va nutq kompetensiyalari, kompyuter savodxonligi, multimedia texnologiyalari, motivatsiya.

Аннотация: В статье рассматриваются вопросы использования современных информационно-коммуникационных технологий при обучении языковым дисциплинам в вузе. Информационная компетентность, ставшая одним из основных показателей качественного обучения, является одним из главных условий современного образования, и ее формирование напрямую зависит от активной деятельности обучающихся в открытой информационно-образовательной среде. Современные информационные технологии позволяют создавать уникальные дидактические условия для разработки и упорядочения индивидуальных образовательных траекторий в обучении языковым дисциплинам.

Использование современных технологий обучения в сочетании с мультимедийными материалами, адаптированными к специфике обучения студентов, а также образовательными ресурсами, имеющимися в свободном доступе, позволяет повысить мотивацию, уровень и качество подготовки студентов при изучении практического курса русского языка. Особую роль в формировании мотивации играют новизна и форма представления получаемой студентом информации, а также их включение в поисковую, творческую деятельность.

Ключевые слова: информационно-коммуникационные технологии, языковая и речевая компетенции, компьютерная грамотность, мультимедийные технологии, мотивация.

Annotation: The article deals with the use of modern information and communication technologies in the teaching of linguistic disciplines in the university. Information competence, which has become one of the main indicators of quality education, is one of the main conditions of modern education, and its formation directly depends on the active activity of students in an open information and educational environment. Modern ICT allows creating unique didactic conditions for the development and ordering of individual educational trajectories in the teaching of linguistic disciplines in the university.

The use of modern teaching technologies in combination with multimedia materials adapted to the specifics of student learning, as well as educational resources that are freely available, can increase the motivation, level and quality of students' preparation when studying the practical course of the Russian language. A special role in the formation of motivation is played by the novelty and form of presentation of the information received by the student, as well as their inclusion in the search, creative activity.

Keywords: information and communication technologies, language and speech competence, computer literacy, multimedia technologies, motivation.

Основной проблемой современного педагогического процесса является дальнейшее совершенствование системы образования в целом, направленное на развитие творческой личности обучающегося. Для целенаправленного и систематического развития интеллекта и творческого мышления обучающихся необходимо применять на занятиях современные педагогические технологии, направленные на активизацию и интенсификацию деятельности обучающихся.

Как показывает практика, ориентация перехода к новой образовательной системе на повышение образования учащихся и его конечный результат вызвала необходимость использования новых технологий обучения и выбора наиболее оптимальных методов и приёмов обучения.

Современное образование, общее и профессиональное, не может быть реализовано в полной мере без использования информационно-коммуникационных технологий (ИКТ).

Интернет, мобильные и спутниковые средства связи, так же, как и традиционные средства массовой коммуникации, помогают совершенствовать общие и профессиональные компетенции, способствуют развитию всех видов речевой деятельности на иностранном языке. Новые виды коммуникационного взаимодействия приводят к новым формам устного общения. Международная интеграция и коммуникативная адаптация к современному мультинациональному и межкультурному пространству оказывают большое влияние на формирование умений устной речи на иностранном языке. [8]

В настоящее время основным источником информации почти во всех областях знаний является Интернет. Современные устройства, такие как смартфоны, планшетные компьютеры, нетбуки, ноутбуки, использующие технологию Wi-Fi, предоставляют большие возможности применения их почти на всех занятиях в любой аудитории.

Большое значение для развития умений устной речи имеет моделирование в учебном процессе профессионально ориентированной среды, имитирующей реальные ситуации общения. Такое моделирование предполагает, что способы решения задач проблемного характера определяются самими студентами на основе их интересов, индивидуальных

особенностей, потребностей, мотивов, способностей, т.е. устанавливается непосредственная связь учебного материала с жизненным опытом обучающихся [1].

Сегодня компьютеризация охватила практически все сферы человеческой деятельности. Уже не представляется возможным получение образования без информационно коммуникационных технологий.

При анализе целесообразности использования ИКТ в учебном процессе можно выявить следующие его дидактические возможности:

- расширение самостоятельной творческой деятельности учащихся, особенно при исследовании и систематизации явлений изучаемого иностранного языка;
- привитие навыков самоконтроля, самостоятельного исправления ошибок;
- развитие познавательных способностей;
- быстрое освоение материала;
- интегрированное обучение предмету;
- развитие мотивации учащихся;
- экономия времени;

Эти дидактические возможности доказывают эффективность и актуальности использования ИКТ в преподавании гуманитарных дисциплин, в том числе русского языка.

К сожалению, при использовании мультимедийных технологий на занятии возникает ряд проблем:

- недостаточное обеспечение учебного процесса компьютерами;
- отсутствие программного обеспечения в условиях поликультурного региона.

И все же доказано, что использование мультимедийных технологий на занятиях дает высокие результаты, т.к. оно способствует развитию творческих, исследовательских способностей обучающихся, повышает их активность. Следует отметить, что далеко не все темы и аспекты работы могут быть компьютеризированы. Прежде всего использование мультимедийных технологий возможно для:

- объяснения нового материала с целью его максимального усвоения;
- оптимального закрепления пройденного материала;
- формирования прочных орфографических и пунктуационных умений и навыков;
- обогащение словарного запаса;
- овладение нормами литературного языка;
- изучение лингвистических и литературоведческих терминов;
- улучшения контроля знаний ученика (проведения уроков-зачетов, тестирования по различным темам, словарно-орфографических и контрольных диктантов и др.).

В условиях многоязычия Республики Узбекистан использование инновационных технологий в преподавании русского языка обусловлено современными направлениями в обучении, предполагающими:

- коммуникативную направленность (работа с текстом, анализ актуальных проблем, усвоение правил в процессе выполнения заданий);
- текстоцентрический подход (многофункциональность текста: текст как эталон для собственного текста; текст как база для организации бесед и дискуссий);
- функциональный подход (что, как и зачем использовано в тексте?);

— культуроведческий подход (расширение знаний учащихся с помощью текстового материала о русской и национальной культуре, о взаимосвязи русской культуры с культурами других народов);

— сопоставление (один из важнейших приемов методики обучения в условиях многоязычия).

Эффективное использование мультимедийных технологий возможно на этапе объяснения новой темы. Например, при изучении любой темы можно использовать гипертекстовый документ, созданный в WORDe. Берется чистый лист, на нем набирается текст предложения, с которого делаются ссылки на другие документы.

Если учащийся не помнит, например, что такое «классицизм», то в ссылке даются сведения об этом литературном направлении. Такая работа хороша тем, что можно самостоятельно при помощи компьютера повторять практически весь материал. На уроке создается благоприятный психологический климат, так как студент не комплексует из-за незнания темы, а самостоятельно добывает знания при помощи гипертекста.

Наблюдения показывают, что после занятий в компьютерном классе учащиеся начинают серьезнее относиться к орфографии и пунктуации, они убеждаются, что орфографическая или пунктуационная безграмотность создает реальные помехи для коммуникации.

При обучении русскому языку как иностранному широко используются возможности MicrosoftOffice 2010, которые позволяют создавать текстовые, графические и иные материалы, способствующие наглядному и эффективному введению и закреплению новой лексики и грамматики.

Используемые на начальном этапе обучения студентов национальных групп специализированные обучающие программы, такие как, например, «Праздники России», «Третьяковская галерея», «Путешествие по Петербургу», «Пушкин и Навои: две гении поэзии», «Лев Толстой – великий русский писатель» и другие ориентированы на формирование лингвострановедческой компетенции. [3]

Использование информационных технологий позволяет достаточно успешно решать следующие задачи:

- обеспечить индивидуальную траекторию обучения студентов;
- обеспечить визуализацию учебной информации;
- обеспечить в реальном режиме времени самоконтроль за степенью и качеством усвоения учебного материала;
- усилить мотивацию обучения за счёт новизны работы с современной вычислительной техникой и обучающими программами.

Использование возможностей существующего программного обеспечения, в частности Microsoft Power Point, позволяет создавать красочно оформленные и информационно насыщенные презентации, которые обладают широкими возможностями при обучении студентов фонетике, лексике русского языка и по осмыслению грамматических правил. Представление лексико-грамматических правил в виде текстового материала, а также таблиц и схем способствуют повышению эффективности усвоения теоретического материала, активизации мыслительной деятельности, развитию логического и абстрактного мышления студентов. [4]

Проанализировав опыт применения ИКТ при проведении учебных занятий по русскому языку (как иностранному) и самостоятельной работе студентов, можно сделать вывод о том, что ИКТ ускоряют процесс обучения, способствуют повышению интереса студентов к изучаемой дисциплине, улучшают качество усвоения материала, позволяют индивидуализировать процесс обучения и обеспечивают возможность объективности оценки. Увеличивается объём и улучшается качество знаний студентов, повышается мотивация к обучению, процесс обучения становится более наглядным и понятным.

Применение инновационных технологий в языковом обучении делает практическое занятие привлекательным и по-настоящему современным. Можно сделать вывод, что такая организация учебной деятельности играет большую роль в развитии коммуникативной компетенции обучающихся, способствует развитию их речевых возможностей, развивает интерес к изучаемому языку и культуре.

Использованная литература:

1. Сокол И.А. Проект как метод реализации коммуникативного подхода в обучении иностранному языку // Иностранные языки в школе. – 2008. – № 1. – С. 16–21.
2. Кудрявцев В.Т. Проблемное обучение: истоки, сущность, перспективы. –М.: Знание, 1991.
3. Ахмедова М.Х., Хосилова Ф.Р. Основные направления оптимизации обучения русскому языку в вузе на базе ИКТ // Ta'lim, fan va innovatsiya. 2020. - №3. С. 80-85.
4. Буримская Д. В. Обучение студентов иностранному языку на базе ИКТ // Информационное общество. 2017. № 6. С. 61–67
5. Бовтенко М.А. Профессиональная информационно-коммуникационная компетенция преподавателя иностранного языка. Новосибирск: Изд-во НГТУ, 2005
6. Воронкова О.Б. Информационные технологии в образовании. Ростов-на-Дону: Феникс, 2010
7. Полат Е.С., М.Ю. Бухаркина, и др. Новые педагогические и информационные технологии в системе образования. М., Академия, 2000.
8. Тарамова, Э.А. Проблемы и перспективы использования ИКТ в высшей школе // Актуальные задачи педагогики: материалы VI Междунар. науч. конф. (г. Чита, январь 2015 г.). – Чита: Издательство Молодой ученый, 2015. С. 156.

ИСПОЛЬЗОВАНИЕ СОВРЕМЕННЫХ ИКТ В ПРОЦЕССЕ ОБУЧЕНИЯ ЯЗЫКАМ

Н.Ш. Иногамова

Стар. преп. Национального
Университета Узбекистана

Annotatsiya: Mazkur maqolada til o'qitish jarayonida zamonaviy axborot kommunikativ texnologiyalaridan foydalanish hamda bo'lajak mutaxassislar kommunikativ kompetensiyasini shakllantirishning ahamiyati haqida fikr yuritiladi. Shuningdek, maqolada multimedia mashg'ulotlari mavzulari ko'lami va ularni amalga oshirish shakllari haqida tavsiyalar beriladi.

Kalit so'zlar: zamonaviy texnologiyalar, kommunikativ texnologiyalar, didaktik imkoniyatlar, multimediali texnologiyalar, orfografik va punktatsion bilim va ko'nikmalar, tekstosentrik yondashuv.

Annotation: This article discusses the use of modern ICTs in the process of learning languages and the role of shaping the communicative competence of future specialists. The article also recommends the subject of multimedia classes and forms of their conduct.

Keywords: modern technology, communication technologies, didactic capabilities, multimedia technologies, spelling and punctuation skills, text-centric approach.

Аннотация: В данной статье рассматривается использование современных ИКТ в процессе обучения языкам и роль формирования коммуникативной компетенции будущих специалистов. Также в статье рекомендуется тематика мультимедийных занятий и формы их проведения.

Ключевые слова: современная технология, коммуникационные технологии, дидактические возможности, мультимедийные технологии, орфографические и пунктуационные умения и навыки, текстоцентрический подход.

В стремительно развивающийся XXI веке преподаватель должен владеть современными технологиями и пользоваться ими на занятиях в вузе. С помощью таких технологий достигается овладение знаниями, умениями и навыками, формирование мировоззрения студентов и развитие их способностей. Сейчас уже все понимают, что современные информационные и коммуникационные технологии обладают колоссальными обучающими возможностями и не менее впечатляющими услугами.

Использование информационных технологий является одной из актуальных проблем современной методики преподавания языковых дисциплин. Применение информационных технологий считается необходимым на уроках, и мотивируется это тем, что они способствуют совершенствованию практических умений и навыков, позволяют эффективно организовать самостоятельную работу. С помощью ИКТ индивидуализируется процесс обучения, повышается у учащихся интерес к занятиям, активизируется их познавательная деятельность, урок становится современным, привлекательным.

При анализе целесообразности использования компьютера в учебном процесса можно выявить следующие его дидактические возможности:

- развитие познавательных способностей;
- быстрое освоение материала;
- интегрированное обучение предмету;
- развитие мотивации учащихся;
- экономия времени;

Эти дидактические возможности доказывают эффективность и актуальности использование компьютера в преподавании гуманитарных дисциплин.

К сожалению, при использовании мультимедийных технологий на занятии возникает ряд проблем:

- недостаточное обеспечение учебного процесса компьютерами;
- отсутствие программного обеспечения в условиях поликультурного региона.

И все же доказано, что использование мультимедийных технологий на занятиях дает высокие результаты, т.к. оно способствует развитию творческих, исследовательских способностей обучающихся, повышает их активность. Следует отметить, что далеко не все темы и аспекты работы могут быть компьютеризированы. Прежде всего использование мультимедийных технологий возможно для:

- объяснения нового материала с целью его максимального усвоения;
- оптимального закрепления пройденного материала;
- формирования прочных орфографических и пунктуационных умений и навыков;
- обогащения словарного запаса;
- овладения нормами литературного языка;
- изучения лингвистических и литературоведческих терминов;
- улучшения контроля знаний студента (проведения уроков-зачетов, тестирования по различным темам, словарно-орфографических и контрольных диктантов и др.).

В условиях многоязычия Республики Узбекистан использование инновационных технологий в преподавании узбекского (русского, английского) языка обусловлено современными направлениями в обучении, предполагающими:

- коммуникативную направленность (работа с текстом, анализ актуальных проблем, усвоение правил в процессе выполнения заданий);
- текстоцентрический подход (многофункциональность текста: текст как эталон для собственного текста; текст как база для организации бесед и дискуссий);
- функциональный подход (что, как и зачем использовано в тексте?);
- культуроведческий подход (расширение знаний учащихся с помощью текстового материала о национальной культуре, о взаимосвязи русской культуры с культурой узбекского народа);
- сопоставление (один из важнейших приемов методики обучения в условиях многоязычия).

Эффективное использование мультимедийных технологий возможно на этапе объяснения новой темы. Например, при изучении любой темы можно использовать гипертекстовый документ, созданный в WORDe. Берется чистый лист, на нем набирается текст предложения, с которого делаются ссылки на другие документы.

Наблюдения показывают, что после занятий в компьютерном классе учащиеся начинают серьезнее относиться к орфографии и пунктуации, они убеждаются, что

орфографическая или пунктуационная безграмотность создает реальные помехи для коммуникации.

Итак, компьютер является хорошим помощником наряду с другими средствами обучения таким разделам языкознания, как:

— словообразование: анализ структуры слов; «разбор слова» (найти часть слова); из данных морфем образовать новые слова; составить слова по модели; проанализировать этимологию слов;

— морфология: «части речи» (найти заданные части речи в произведении);

— лексикология: «Фразеология» (найти аналогичные варианты фразеологизмов на родном языке или их синонимы);

-- синтаксис: «Сложное предложение» (из данных частей собрать сложное предложение – ССП,СПП).

На завершающем этапе изучения проблемы, рекомендуются следующие темы-проекты:

— «Особенности перевода с русского языка на родной»;

— дискуссия «Экология языка»;

— конкурс «Рецепты национальной кухни»;

— дискуссия «Современная музыка»;

— «Творческая биография любимого деятеля культуры»;

— сочинение-рассуждение «Современная молодежь и ее речь»;

Мультимедийные технологии удобны своей универсальностью. Они могут быть использованы учащимися как для самоконтроля, так и для контроля со стороны преподавателя.

Таким образом, отметим выявленные на практике положительные результаты данной технологии:

- совершенствует навыки и умения говорения и письма;

- развивает творческую способность учащихся;

- учит экономить время, организует собранность и внимательность, развивает мыслительную деятельность;

- из-за сжатости времени возникает «мозговая атака», которая реализуется за счет интеллектуального уровня.

Из всего многообразия педагогических технологий мы в большей степени остановили свое внимание на том, который может быть достаточно органично и легко интегрироваться с традиционным обучением. Именно такие технологии являются основными векторами развивающего обучения. Здесь налицо качественно иной подход преподавателя, не являющийся способом разукрасить занятие, доставить учащимся удовольствие от использованных игровых приемов, групповых форм работы, частой смены деятельности. Это технологии с совершенно четкой структурой, имеющие в своей основе развивающие и воспитательные цели.

Возможность работы в компьютерных сетях открывает перед преподавателями и студентами новые перспективы для творческого развития личности, накопления опыта, самостоятельного получения знаний, зачастую более полных, а иногда и отличных от предоставляемых в печатно-изданных учебниках и справочных материалах.

Сеть интернета создает уникальную возможность для студентов, изучающих язык пользоваться аутентичными текстами, слушать и общаться с носителями языка, то есть он создает естественную языковую среду.

Социальные сети тоже могут использоваться в образовательных целях. Студенты, создавая свои личные странички на изучаемом языке, развивают письменную речь, получают возможность делиться полезной информацией, вступать в профессиональные сообщества и группы по интересам.

Эффективность изучения языка заключается в том, как часто мы на нем общаемся. Использование ИКТ при обучении языкам помогает восполнить пробел в общении, практически применять и закреплять полученные на традиционном занятии теоретические знания.

В заключении отметим, что использование современных ИКТ вместе с новыми педагогическими разработками значительно повышают эффективность обучения и ее роль в формировании современного образа мышления будущих специалистов. Обучение языку – именно та область, где использование информационно-коммуникативных технологий может принципиально изменить и методы работы, и, самое главное, её результаты.

Использованная литература:

1. Азизходжаева Н.Н. Таълим технологиялари ва педагогик маҳорат - Т., 2006.
2. Беспалько В.П. Педагогика и прогрессивные технологии обучения. - М., 1995.-127с.
3. Сысоев П.В. Информационные и коммуникационные технологии в лингвистическом образовании: Учебное пособие. М.: Либроком, 2013.
4. Сайидахмедов Н.С. Педагогик маҳорат ва таълим технологиялари. – Т.: ЎзМУ, 2003
5. Волошина Н.Н. «Творческая педагогика», №3 2001г.
6. <http://nsportal.ru/skola/>

TIL O'RGANISHDA KYTO' VA MYTO'DAN FOYDALANISH

Saodat Boysariyeva Baxtiyor qizi,
TerDU o'qituvchisi

Annotatsiya: Ushbu maqolada til o'rganishda kompyuter texnologiyalari va mobil qurilmalardan foydalanish va uning nazariy asoslarni tadqiq qiluvchi KYTO' va MYTO' ga oid tadqiqotchilarning fikrlarini tahlil qilingan.

Kalit so'zlar: elektron lug'atlar, KYTO' (CALL), MYTO' (MALL), m-o'rganish (m-learning), o'quv lug'atlari, leksikografiya, til ta'limi

Аннотация: В данной статье анализируются мнения исследователей относительно использования компьютерных технологий и мобильных устройств в изучении языка и его теоретических основ, KYTO и MYTO.

Ключевые слова: электронные словари, KYTO, MYTO, m-learning, учебные словари, лексикография, языковое образование.

Annotation: This article analyzes the opinions of researchers regarding the use of computer technology and mobile devices in language learning and its theoretical foundations, KYTO and MYTO.

Keywords: electronic dictionaries, KYTO (CALL), MYTO (MALL), m-learning, educational dictionaries, lexicography, language education.

Elektron lug'atlar til o'rganishda samarali vosita ekanligi haqida S. Atkins[1], H. Nesi[2], H. Bejoint[3] kabi qator tadqiqotchilar o'z mulohazalarini berishgan.

Daslabki elektron o'quv lug'atlari:

- The Longman Interactive Dictionary (Longman interaktiv lug'ati) (1993)
- The Elektronik Oxford Word-power Dictionay (Elektron Oksford so'z kuchi lug'ati) (1994)
- Collins Cobuild on CD-ROM (CD-ROM dagi Kollinz Kobild) (1995)

Maskur lug'atlar bosma va elektron o'quv lug'atlari orasida ko'prikdir. Nesining fikricha ulardan kompyuter ilovalari uchun to'g'ridan to'g'ri havolalar olish, qidiruv asosida izlash va lug'at foydalanuvchilari bilan muloqot o'rnatish imkonini beradi. O'sha davrdayoq Sue Atkins leksikografiyaga texnologiyalarni jalb etish chet tili o'rganuvchilarning ehtiyojlarini ancha qondirishini bashorat qilgan edi. Uningcha, elektron lug'atlar foydalanuvchilarga alifbo tartibisiz qidiruv, ma'lumot taqdim qiloyotganda grafiklardan foydalanish, katta hajmdagi korpuslar bilan bog'lana olish kabi imkoniyatlarni yaratadi. Atkinsning keltirgan fikrlari asosan ikki tilli lug'atlar haqida edi.

Keyinchalik leksikografiya va kompyuter texnologiyalarining birlashuvi fanda yangi soha paydo bo'lishiga turtki bo'ldi. Jahon tilshunosligida bu soha CALL (computer assisted language learning) - KYTO' (kompyuter yordamida til o'rganish) nomi bilan yuritiladi. KYTO' kompyuter orqali til o'qitish va o'rganishni yangi texnologiyalar va media vositasida amalga oshiradi. Ko'plab adabiyotlarda KYTO' (CALL) ning paydo bo'lishini o'tgan asrning 60-yillariga to'g'ri kelishini aytib o'tilgan[4]. O'sha yillarda daslabki tajribalardan biri 1967- yilda kompyuter o'quv

korporatsiyasini tashkil etish maqsadida Atkinson va Suppes tomonidan tajribadan o'tkazilgan. Keyingi tajriba grammatik tarjima usuliga asoslangan, lug'at va grammatik mashqlar bilan boyitilgan PLATO dasturi edi.

Mutaxassislar KYTO' (CALL) ni 3 bosqichga ajratadilar:

1. Strukturaviy KYTO' (1960-1980)
2. Kommunikativ KYTO' (1980-1990)
3. Integrativ KYTO' (1990 ~)

KYTO' ning har 3 bosqichi bir-birini rad etmaydi, bir birini to'ldirib boradi. Kompyuter yordamida til o'rganishdagi keng ko'lamlı ishlarning aksariyati chet tilini o'rgaishga yo'naltirilgan bo'lib, uning ona tilini o'qitish bilan bog'liq jihatlarni ko'p o'rganilmagan. Hozirda KYTO' leksikografiyaning ajralmas qismi.

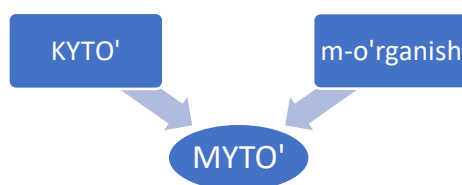
Til o'qitishning keyingi bosqichi MYTO' (mobil yordamida til o'rganish)— MALL (Mobile asisted language learning)dir. MYTO' orqali til o'rganish atamasi 1-marta 2006- yilda Chinney tomonidan qo'llanila boshlagan. O'sha vaqtda MYTO' KYTO'ning tarkibiy qismi edi. Keyinchalik Kukulska va Shield 2008- yilda MYTO' va KYTO' o'rtasidagi farqni ochib berishgan.

Internet va mobil qurilmalardan keng foydalanish KYTO'dan MYTO'ning ajralib chiqishiga olib keladi. MYTO'ning KYTO'dan asosiy farqli jihatlari quyidagi o'rinlarda ko'zga tashlanadi:

- portarativlik yoki mobillik
- ijtimoiy aloqalarning kengligi
- yuqori darajadagi individuallik

KYTO' ushbu xususiyatlarni talab darajasida ta'minlab bera olmaydi[5].

MYTO' hozirda ommalashgan mobil o'rganish — m-learning (mobile learning)ning rivojlanib borayotgan sohasidir.



Mobil yordamida til o'rganish **KYTO'** hamda **m-o'rganish** integrallashuvi natijasida paydo bo'lgan yo'nalish bo'lib, uni rivojlantirish va muammolarni bartaraf etish maqsadida ta'lim sohasida Yevropada MLEARN seminarlari, MobilLearn, MLang kabi loyihalar tashkil etilmoqda. Jek Butson[6].

1994-2012- yillarda MYTO' sohasida 345 ta ilmiy nashrlarni mavjudligini keltirib o'tgan, Chinney, El-Husayn, C. Houser, E. Klopfer, K. Squire, H Jenkins, A. Kukulsha- Hulme, H. Li, M. Levy, Kennedi, M. Sharoles va b.q. tadqiqotchilar o'z ishlari bilan ushbu sohaning nazariy jihatlarni yoritib borishmoqda. Mobil qurilmalar yordami til o'rganish haqida foydalanuvchilar(talaba va o'quvchilar)ning fikrlari ijobiy. Hozirda mobil qurilmalaridan o'quv maqsadida foydalanish mumkin. Ba'zi mutaxassislarning fikriga ko'ra MYTO' grammatikani o'rganishnii osonlashtiradi. Tadqiqotchilar o'qituvchi va pedogoglarni MYTO'ga jalb etish sohani yana rivojlantiradi deb hisoblashadi.

Mamlakatimizda olib borilayotgan til islihotlari ona tilini zamonaviy vositalar yordamida o'qitish, til ta'limida jahon tajribasidan o'tgan, foydalanuvchilarni o'ziga jalb etuvchi yangi sohalarining yutuqlaridan unumli foydalanish til ta'limida va leksikografiyamizda yangi sahifaning ochilishi deb baholash mumkin.

Foydalanilgan adabiyotlar:

1. B.T. Sue Atkins (ed.), 1998, *Using Dictionaries: Studies of Dictionary use by Language Learners and Translators*. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.14734192.2000.tb00146.x/>
2. H. Nesi Dictionaries in electronic form /The Oxford History of English Lexicography (book pp.458-478) Oxford University Press/ January 2009
3. H. Bejoint, The Lexicography of English [International Journal of Lexicography](https://doi.org/10.1093/ijl/ecq042) August 2010 /24(3):359-363/ DOI:[10.1093/ijl/ecq042](https://doi.org/10.1093/ijl/ecq042)
4. MS Xavier Pradheep Singh / The History and the Current Status of Computer Assisted Language Learning/ Journal of English Language Teaching (India), Volume LVII Number 5 September–October, 2015, 25-32
5. Burston, J. (2013). Mobile-assisted language learning: A selected annotated bibliography of implementation studies 1994–2012. *Language Learning & Technology*, 17(3),157–224 <http://lt.msu.edu/issues/october2013/burston.pdf>
6. Burston, J. (2013). Mobile-assisted language learning: A selected annotated bibliography of implementation studies 1994–2012. *Language Learning & Technology*, 17(3),157–224 <http://lt.msu.edu/issues/october2013/burston.pdf>

**INTELLEKTUAL TAHLIL ASOSIDA CHET TILI BILIMINI TEKSHIRISHNI
MODELLASHTIRISH****Z.Sh. Abdullayeva**

f.-m. fanlari falsafa doktori, dotsent

Muhammad al-Xorazmiynomidagi

Toshkent Axborot Texnologiyalari Universiteti

G.A. Shihnazarova

Muhammad al-Xorazmiy nomidagi Toshkent Axborot

Texnologiyalari Universiteti tayanch doktoranti

Annotatsiya: Maqolada ingliz tili bilish darajasini intellektual tahlil qilish asosida baholashni modellashtirish to'g'risida fikr yuritilgan. Testlash, ta'lim muhiti, natijalarni aniqlash va saqlash kabi modullarni qamrab oluvchi bilimlarni baholash modeli ishlab chiqilgan, har bir modulning ta'rifi keltirilgan va ingliz tilini o'qitish jarayoni ko'rib chiqilgan.

Калит сўзлар: o'qitish tizimlari, modul, bilimni baholash, modellshtirish.

Аннотация: В данной статье приведено моделирование процесса проверки знаний на основе интеллектуального анализа данных при изучении английского языка. Разработана модель проверки знаний, которая включает следующие модули: тестирования, сохранения результатов и получения данных. Приведено описание каждого модуля и рассматривается процесс обучения английского языка.

Ключевые слова: обучающие системы, модуль, оценка знаний, моделирование.

Annotation: This article shows simulation of the process of examination based on data mining techniques in the study of English language. The model of the knowledge test, which includes the following modules: testing, store the results and data acquisition. The description of each module and covers the process of learning English.

Keywords: English language, module, knowledge assessment, modeling.

Bugungi kunda tezkorlik bilan rivojlanayotgan axborot kommunikatsiya texnologiyalari imkoniyatlaridan to'laonli foydalanib, ta'lim jarayoniga yangicha yondashish va uni tashkil etish muhim vazafalardan biri hisoblanadi. Ayniqsa, AKT orqali chet tillarini o'qitish va o'rganish sohasiga qaratilgan yuksak e'tibor buning yaqqol isbotidir.

O'zbekiston Respublikasi Prezidentining "Chet tillarni o'rganish tizimini yanada takomillashtirish chora-tadbirlari to'g'risida"gi 2012-yil 10-dekabrdagi PQ-1875-son qarorini bajarish yuzasidan Vazirlar Mahkamasi qaroriga binoan uzluksiz ta'lim tizimining chet tillar bo'yicha davlat ta'lim standarti "Chet tillari bo'yicha ta'limning barcha bosqichlari bitiruvchilarining tayyorgarlik darajasiga qo'yiladigan talablar" 2013-2014 o'quv yilidan boshlab bosqichma-bosqich amaliyotga joriy etildi.

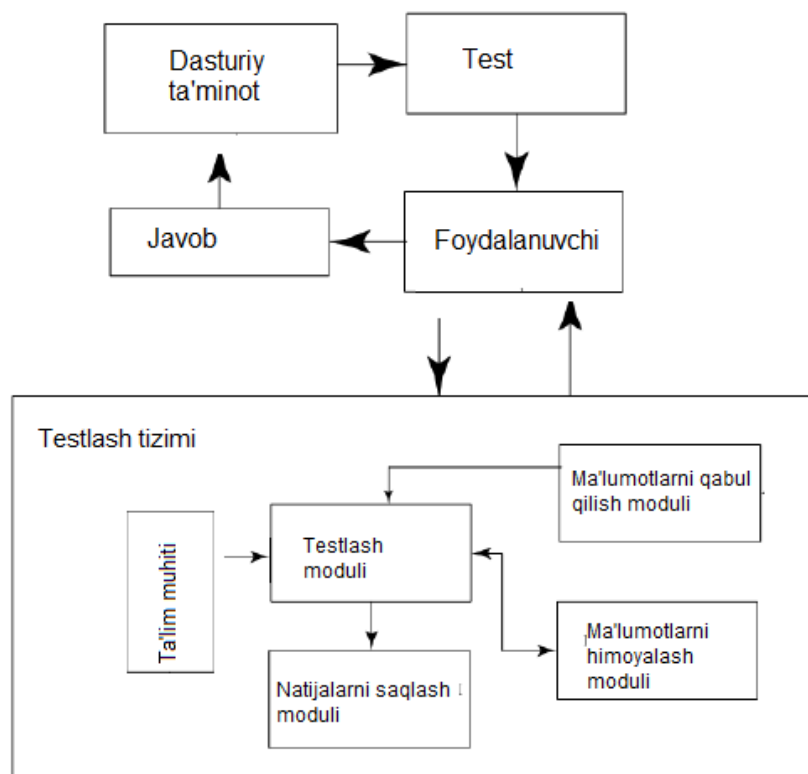
Chet tillarini o'qitishda ta'lim va pedagogik jarayonlarga Internet hamda uning katta imkoniyatlaridan foydalanilishi mazkur qarorning yanada istiqbolliligini ta'minlaydi. Shunga

binoan bir qancha ingliz tilini o'rgatuvchi, bilimni baholovchi har xil web saytlar milliy tarmog'imizda o'rin olib kelmoqda.

Bugungi kunda bir qancha kompyuter orqali bilim beruvchi va tekshiruvchi intellektual ta'lim tizimlari joriy qilingan bo'lib AOC-BY3, АДОНИС, МастерТест, TestMaker ta'lim tizimlari shular jumlasidandir. Quyida tizimning ish olib borish tartibi ko'rsatilingan. Foydalanuvchiga dastlab o'rganilishi kerak bo'lgan ma'lumot taqdim qilinadi, so'ngra shu mavzuga oid bir qancha savol berib tekshiriladi, ya'ni taqdim qilingan axborot necha foiz o'rganib chiqildi. Agar foydalanuvchi noto'g'ri javob bersa, qo'shimcha ma'lumotlar berilib, qayta tekshiruv o'tkaziladi, to'g'ri javob berilsa keyingi darajaga o'tkaziladi va shu tariqa bilim o'rgatish va baholash davom qiladi.

Ingliz tilini o'rganishga talab oshayotgan ekan, shunga mos ravishda chet tillarni o'rgatish va bilimni baholovchi dasturiy tizimlar ortib bormoqda.

Ushbu taqdim qilinayotgan model o'zining intellektual tarzda baholash va shu bilan birga o'z navbatida ingliz tilini o'rgatish imkoniyati mavjud ekanligi bilan boshqa tizimlardan farq qiladi. Quyida foydalanuvchi va testlash tizimining o'zaro bog'langan ketma ketlik ifoda etilgan.



1-rasm. Ma'lumotlarni intellektual tahlil qilish asosida bilimlarni tekshirish jarayonining modeli

M1: *Ro'yxatdan o'tish.* Har bir foydalanuvchi o'zining ingliz tili bilimni sinab ko'rishdan oldin ro'yxatdan o'tishi zarur.

M2: *Testni moduli.* Tizimga avtorizatsiya qilingandan keyin foydalanuvchida tanlash imkoniyati mavjud bo'lib:

- a) Ingliz tilini o'rganishda ma'lumotlarni intellektual taxlil qilish asosida bilimlarni tekshirish;

b) Ingliz tiliga oid mavzular va unda ma'lumotlar bilan tanishib chiqib, so'ngra testni boshlash;

M3: *Testning daraja bo'yicha o'zgarishi*. Foydalanuvchi tomonidan belgilangan javobga mos tarzda test tizimining qiyinchilik darajasi o'zgarib boradi. Ya'ni agar to'g'ri javob berilsa, savol murakkablashadi, mobodo noto'g'ri javob aks etsa aksincha qayta o'zlashtirish maqsadida shu mavzuga oid savollar taqdim qilinadi.

M4: *Ta'lim muhiti moduli*. Ushbu modul o'zida har bir mavzuga oid ma'lumotlarni to'plamini qamrab oladi. Savolga noto'g'ri javob berilganda tizim avtomatik tarzda test topshirayotgan foydalanuvchiga mos ravishda ma'lumot taqdim qiladi, so'ngra qay darajada o'zlashtirilganligini aniqlash maqsadida qayta testdan o'tkaziladi.

M5: *Natijani aniqlash moduli*. Taqdim qilingan savollarga javob berilgan keyin tizim avtomatik tarzda natijani e'lon qiladi va foydalanuvchining ingliz tilida bilimini aniqlanadi, so'ngra zaruriy ma'lumotlar, ya'ni uning kamchiliklari va o'rganilishi zarur bo'lgan ma'lumotlar taqdim qilinadi.

M6: *Natijani saqlash moduli*. Foydalanuvchining bilimi aniqlangan keyin, ro'yxatdan o'tilgan paytdagi shaxsiy ma'lumotlari saqlanuvchi jadvaliga ega bo'lgan MB da saqlanadi.

Foydalanuvchilarni ingliz tilini bilish darajasini, bilimlar bazasiga tayangan holda tekshirish va natijalarga bog'liq tarzda yo'nalish berish, agar to'g'ri javoblar kam miqdorni tashkil qilsa, mavzuga oid bo'lgan boshqa ma'lumotlarni ham qayta o'rganish, o'rta miqdor holati kuzatilsa, mavzuni boshdan o'rganib chiqish, yuqori natijaga erishsa, keyingi mavzuni o'rganish kabi vazifalarni taklif qilishga qaratilgan.

Foydalanilgan adabiyotlar:

1. Башмаков А.И., Башмаков И.А. Разработка компьютерных учебников и обучающих систем. – М.: Филинь. – 2003. – 430с.
2. Shixnazarova, G. A., & Gulyamova, D. R. (2016). Modeling of the process of examination data mining techniques in the study of English language. *Информатика: проблемы, методология, технологии*, 900.
3. Каримова, В. А., & Шихназарова, Г. А. (2015). Инглиз тилини ўрганишда билимларни текшириш жараёнини моделлаштириш. *Современное образование (Узбекистан)*, (5), 19-23.
4. Karimova V.A., Shihnazarova G.A., Muhitdinova M.L. Simulation of the process of examination in the study of English language, UDC 004.384.4:37 TUIT BULLETIN 3(35)/2015 <http://jurnal.tuit.uz>, 27-31

3-SHO'BA

NLPDA TIL VA NUTQ TAHLILI (LINGVISTIK ANALIZATOR: MORFOLOGIK, SINTAKTIK VA SEMANTIK TAHLIL, NUTQ ANALIZI VA SINTEZI)

TURKIY TILLAR O'RTASIDA TABIIY TILLARNI QAYTA ISHLASH MODEL (O'ZBEK TILI MISOLIDA)

Nazirova Elmira

t.f.d., professor

Muhammad al-Xorazmiy nomidagi TATU

Abidova Shaxnoza Baxodirovna

t.f.f.d., PhD,

Muhammad al-Xorazmiy nomidagi TATU

Uzakova Ma'mura

Doktorant, Muhammad al-Xorazmiy nomidagi TATU

Annotatsiya: Maqolada bugungi kunning dolzarb mavzularidan biri bo'lgan kompyuter lingvistikasini rivojlantirishda NLP dan foydalanish va uning imkoniyatlari, elektron tarjimalarni yaratishda NLP ning o'rni yoritilgan.

Kalit so'zlar: kompyuter lingvistikasi, NLP, elektron tarjima, matematik model.

Bugungi globallashuv davrida insonlar o'rtasida katta hajmli ma'lumotlar bilan almashish, matnli, ovozli, tasvirli va video ma'lumotlarni yuborish va qabul qilish ularni qayta ishlash jarayonlari oshib bormoqda. Bu jarayonda tabiiy tillarni qayta ishlash (NLP) keng imkoniyatlar beradi.

Tabiiy tilni qayta ishlash (NLP) - bu kompyuterlarga tabiiy tillardagi matnlarni tushunish, yaratish va qayta ishlash imkonini beruvchi sun'iy intellektning (SI) yo'nalishlaridan biridir. Tabiiy tildagi matnni qayta ishlash ma'lumotlar so'rovi uchun tabiiy til matnlari yoki og'zaki nutqlardan foydalanish imkonini beradi.

Kompyuter lingvistikasi - lingvistik qonuniyatlarni tavsiflash uchun matematik modellardan foydalanishni amalga oshiradi. Kompyuter lingvistikasini ikkita katta qismga bo'lish mumkin.

Birinchi qism lingvistik tadqiqotlarda kompyuter texnologiyalaridan foydalanish usullarini aniqlash uchun ma'lum matematik usullardan foydalanishni o'rganadi.

Ikkinchi qism lingvistik muammolarni hal qilishning matematik modellarini yaratish va ushbu modellar asosida ishlaydigan dasturlarni ishlab chiqishda foydalaniladi. Kompyuter lingvistikasining bu qismi tabiiy til matnlarini qayta ishlash tizimlarini ishlab chiqish bilan shug'ullanadigan sun'iy intellekt bo'limi bilan chambarchas bog'liq.

Tabiiy tillarga qayta ishlov berish bilan bir qator olimlar va tadqiqotchilar ilmiy izlanishlar olib borganlar va hozirda ham tadqiqotlar olib borilmoqda. Xorijiy olimlardan V.Yu.Yurgel o'zining ilmiy tadqiqotlarida "Tabiiy tillarni qayta ishlash murakkabligi" bo'yicha ishlar olib

borgan [V.Yu.Yurgel, 2019: 13]. Dou Z., Wang X., Shi Sh., Tu Zh. olimlar ham NLP bo'yicha bir qator tadqiqot ishlarini amalga oshirganlar [Dou Z., Wang X., Shi Sh., Tu Zh., 2020: 5]. Jurafsky D., Martin J. o'z tadqiqot ishlarida tabiiy tillarga ishlov berish jarayonlarini ko'rib chiqqanlar [Jurafsky D., Martin J., 2020: 300].

O'zbek olimlaridan tabiiy tillarni qayta ishlash bo'yicha ham bir qator olimlar tadqiqot ishlarini olib bormoqdalar.

O'zbek tilininng korpus lingvistikasini amalga oshirishda hozirgi kunda N.Abduraxmonovanning tadqiqot ishlarini keltirish mumkin. Bugungi kunda ham o'zbek tilining grammatik tahlili, morfologik tahlili so'zlarni bo'laklarga ajratish bo'yicha tadqiqotlar olib bormoqda [Abduraxmonova N., 2021: 250].

Tabiiy tillarni qayta ishlashning qo'llanish sohalari ham bir qancha bo'lib, biz ushbu maqolada mashinali tarjimonda tabiiy tillarni qayta ishlash yo'nalishini ko'rib chiqamiz.

Mashinali tarjimonni amalga oshirishda tabiiy tilni qayta ishlash jarayonining ikkita asosiy usuli qo'llaniladi: sintaksik va semantik tahlil.

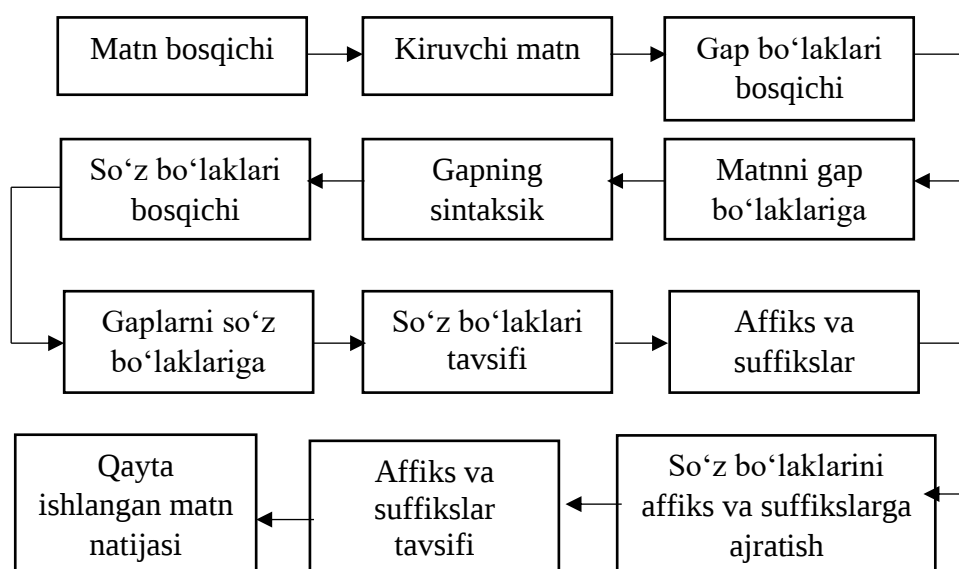
Sintaksik tahlil grammatik aniqlikni ta'minlash uchun so'zlarni joylashtirishga imkon beradi. Aniq va mukammal tushunish uchun tabiiy tilning grammatik qoidalarga muvofiqligini baholaydi. Sintaksik tahlil yordamida so'z turkumlari ajratib olinadi va grammatik qoidalar orqali ularning ma'nosini bildiradi.

Matnlarning sintaksik tahlili bir qancha usullar yordamida amalga oshiriladi:

- Lemmatizatsiya.m• Morfologik segmentlash. • So'zlarni segmentlash.
- Nutqni teglash qismlari. • Tahlil qilish (Parsing). • Jumlani bo'lish.
- Stemming.

Mashinali tarjimani amalga oshirish jarayonida matnlarni sintaksik tahlilini amalsha oshirish yuqorida keltirilgan usullar yordamida amalga oshiriladi, ya'ni bu matnni qayta ishlashda va tarjima jarayonini amalga oshirishda katta rol o'ynaydi.

Turkiy tillar oilasiga kiruvchi tillar o'rtasida mashinali tarjimani amalga oshirishda tabiiy tillarni qayta ishlashni quyidagi sxema ko'rinishida amalga oshirish mumkin:



1-rasm. Tabiiy tillardagi matnlarni qayta ishlash

Yuqoridagi sxemani tavsiflaydigan bo'lsak:

- matn bosqichi: bu yerda matn yuklanadi.
- gap bo'laklari bosqichi: bu bosqichda kiritilgan matn gap bo'laklariga ajratiladi.
- so'z bo'laklari bosqichi: bu bosqichda gap bo'laklari so'z bo'laklariga ajratiladi.
- affikslar va suffikslar bosqichi: bu bosqichda ko'rilyotgan til aglyutinatív tillar guruhiga kirganligi sababli so'zlar o'zak va affikslarga ajratilgan holda tavsiflanadi.

Bu yerda tabiiy tillarni qayta ishlashning morfologik segmentlash, ya'ni morfologik tahlil usulidan foydalaniladi.

Ushbu sxema turkiy tillar oilasiga hamda aglyutinatív tillar guruhiga kiruvchi o'zbek tilidagi matnlarni, tabiiy tillarni qayta ishlab chiqish sxemasi asosida ishlab chiqildi. Ushbu sxemadan turkiy tillar oilasiga kiruvchi hamda aglyutinatív tillar uchun ham qo'llash mumkin.

Ushbu sxemada tasvirlangan ma'lumotlardan kelib chiqqan holda o'zbek tilida sodda va murakkab so'zlarni hosil qilishning matematik model va algoritmlari ishlab chiqildi.

$$C = \sum_{i=1}^n x_i \oplus \downarrow \sum_{j=1}^m a_j \oplus \downarrow \sum_{k=1}^l f_k \oplus \downarrow \sum_{r=1}^q p_r$$

Bu yerda C – murakkab so'z, $\sum_{i=1}^n x_i$ – har bir so'z x to'plam hisoblanadigan so'z asosi, $\sum_{j=1}^m a_j$ –

so'z yasovchi affikslar, $\sum_{k=1}^l f_k$ – shakl yasovchi affikslar, $\sum_{r=1}^q p_r$ – so'z o'zgartiruvchi affikslar,

bu yerda i, j, k va r – affikslar va so'z uzunligi; $\oplus$ – birlashtirish amalini, $\downarrow$ – “bog'lash” (birlashtirish yoki “bog'lanmaslik” (birlashmaslik) amallarini bildiruvchi belgilar bo'lib, ular tabiiy tillarga ishlov beruvchi “kengayuvchi kirish tili” ning tavsifi keltirilgan maqoladan olingan). Yuqoridagi matematik model asosida biz o'zbek tilida murakkab so'zlarni yaratish ketma-ketligini ko'rib chiqamiz.

$$S_1 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j (y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k (y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r (y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$S_2 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{k=1}^l f_k (y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r (y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$\begin{aligned}
 S_3 &= \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases} \\
 S_4 &= \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases} \\
 S_5 &= \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases} \\
 S_6 &= \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases} \\
 S_7 &= \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{bu yerda } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}
 \end{aligned}$$

Bu yerda $C_1, C_2, C_3, C_4, C_5, C_6, C_7$ – so‘z asosi, $I(z)$ – bitta predikat, har bir z so‘zi ikki yoki undan ortiq so‘z birikmasi sifatida ifodalanadi $y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1}$, bu yerda i – affiks uzunligi [Abidova Sh.B. Nazirova E.Sh., 2021: 4352].

Yuqorida keltirilgan matematik model va algoritmdan turkiy tillar oilasiga kiruvchi aglyutinatativ tillar guruhiga ham so‘zlarni hosil qilishda qo‘llash mumkin.



2-rasm. Ikki tilli elektron tarjima asosiy oynasi

Yuqorida keltirilgan rasmda turkiy tillar oilasiga kiruvchi tillar o'zbek va qoraqalpoq tillarida sodda va murakkab so'zlarni tarjima qiluvchi elektron tarjimaning asosiy oynasi keltirib o'tilgan.

Yuqorida ishlab chiqilga matematik model hamda algoritmdan foydalangan holda turkiy tillar oilasiga kiruvchi tillar o'rtasida ya'ni o'zbek va qoraqalpoq tillari o'rtasida elektron tarjimini amalga oshiruvchi dastur ishlab chiqildi. Hozirgi kunda ushbu model va algoritmlar asosida turkiy tillar oilasiga kiruvchi tillar o'rtasida 3 tilli elektron tarjimon yaratish jarayoni olib borishda asos bo'lib xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Юргел В.Ю. Сложности моделирования естественного языка // Вестник науки и образования. 2019 № 23-1 (77). S.12-14.
2. Dou Z., Wang X., Shi Sh., Tu Zh. Exploiting deep representations for natural language processing // Neurocomputing 2020. Vol. 38 (621). P. 1-7.
3. Jurafsky D., Martin J. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, 2020. – 623 p.
4. Abduraxmonova N. Kompyuter lingvistikasi (darslik). – Toshkent: Nodirabegim, 2021. – 400 b.
5. Abduraxmonova N. O‘zbek tili elektron korpusining kompyuter modellari (monografiya) – Germaniya: Globeedit, 2021. –200 b.
6. Abidova Sh.B. Nazirova E.Sh. Mathematical Model and Algorithm for Calculating Complex Words int Uzbek Language // journal “Solid State Technology”. 2021. Vol. 64. Issue 2. P. 4347-4359.
7. Nazirova E.Sh., Abduraxmonova N.Z., Abidova Sh.B., Uzakova M. O‘zbek tili o‘quv lug‘atlarining elektron platformasini yaratish // O‘ZBEKISTON XOTIN-QIZLAR “OLIMA” UYUSHMASINING 30 YILLIK YUBILEYIGA BAG‘ISHLANGAN Barqaror rivojlanish maqsadlariga erishishda xotin-qizlarning roli xalqaro ilmiy-amaliy anjuman materiallari. 2022. Tom 2. 246-250 b.

МАТНЛИ ХУЖЖАТЛАРНИ ТАСНИФЛАШНИНГ ГИБРИД АЛГОРИТМИНИ ИШЛАБ ЧИҚИШ

Бобоев Лочинбек

Техника фанлари фалсафа доктори,
Япония рақамли университети

Аннотация: Ушбу мақолада ўзбек тилидаги матнли хужжатларга дастлабки ишлов бериш ва таснифлаш алгоритмлари ишлаб чиқилган. Таснифлаш натижалари ЎзА ахборот ресурсларида модел масала сифатида тажрибавий тадқиқот натижалари кўринишида келтирилган. Чуқур ўқитишга асосланган алгоритмлар ўртача 80-90% аниқликни таъминлаган. Гибрид алгоритм эса 93% дан юқори натижа қайд этилган.

Калит сўзлар: матн, таснифлаш, гибрид, cbow, cnn, rnn.

Аннотация: В статье разработаны алгоритмы предварительной обработки и классификации узбекских текстовых документов. Результаты классификации представлены в виде результатов экспериментальных исследований в виде модельной задачи в информационных ресурсах УзА. Алгоритмы, основанные на глубоком обучении, обеспечивали точность в среднем 80-90%. Гибридный алгоритм имеет результат выше 93%.

Ключевые слова: текст, классификация, гибридный, cbow, cnn, rnn.

Annotation: In this article Algorithms for preprocessing and classification of Uzbek text documents are developed. Classification results are presented in the form of experimental research results as a model problem in UzA information resources. Algorithms based on deep learning provided an average accuracy of 80-90%. The hybrid algorithm has a result higher than 93%.

Keywords: text, classification, hybrid, cbow, cnn, rnn.

Матнлар кетма-кетлигини ҳисобга олувчи ёндашув табиий тил матнини таҳлиллашнинг янги босқичи – матнни тушунчавий (мазмун) таҳлиллашни амалга оширади [Бабомуратов, 2020; 192].

Масала ечимига ёндашувнинг бундай кўриниши инсоният томонидан глобал тармоқда ахборот истемоли билан уни ишлаб чиқариш ўртасида катта тафовутнинг мавжудлиги ҳолатини бартараф этиш учун муҳим омил ҳисобланади. Албатта катта ва жуда йирик тизимлар ўзида мазкур ечимни турли кўринишда амалга оширмоқда.

Ҳозиргача сўзларни узлуксиз вектор сифатида ифодалаш кўпгина ишларда амалга оширилган [Хинтон, 1986; 77, Румелхарт, 1986; 533; Элман, 1990; 179]. Электрон хужжатда матн кетма-кетлигини ифодалашнинг узлуксиз сўзлар тўплами (CBOW) ва Skip-gram модели Бенгио томонидан биринчи бор таклиф этилган, бироқ мазкур ёндашув Николаш Миколов бошчилигидаги тадқиқотчилар гуруҳи томонидан ҳам таклиф қилинган [Миколов, 2013; 1, Миколов, 2013; 3111]. Word2vec дастурий воситаси томонидан амалга оширилган сўзлар кетма-кетлиги асосида башоратлаш функциясидан фойдаланишда қулайлиги билан оммалашган.

Кўпгина амалий дастурларда матнни ифодалашда сўзларнинг кетма-кетлиги, жойлашуви ва грамматикасини инобатга олмайдиган сўзлар портфели моделидан фойдаланиш етарли бўлсада, матнни ифодаланишининг кетма-кетлик моделини қўлланиши муҳим эканлигини кўрсатувчи ҳолатлар мавжуд.

Кейинги босқичда нейрон тармоғининг янада мукаммаллаштирилган чуқур ўқитишга асосланган модели ёрдамида матнлар таснифланади.

Кўпгина тадқиқотчилар янги таснифлаш вазибалари учун янада мустаҳкам ва аниқ архитектура ишлаб чиқиш учун стандарт чуқур ўқитиш архитектурасини бирлаштирадilar [Бабомуратов, 2021; 177].

Конволюцион қатламлар ёрдамида бирикмалар даражасидаги белгиларларни ўрганиш учун конволюцион нейрон тармоқ билан LSTM бирлаштирилади.

Гибрид алгоритм қуйидаги қадамлардан иборат:

1-қадам: $D = \{w_1, w_2, \dots, w_n\}$ кириш ҳужжати дастлабки ишлов бериш босқичидан ўтказилади.

2-қадам: Ҳужжатдаги ҳар бир w_i сўзлар векторига трансформация қилинади, $v_i \in \mathbb{R}^d$.

3-қадам: $w_{ii+k-1} = \{w_i, w_{i+1}, \dots, w_{i+k-1}\}$ сўзлар кетма-кетлиги учун мос равишда $v_{ii+k-1} = \{v_i, v_{i+1}, \dots, v_{i+k-1}\}$ векторлар шакллантирилади.

4-қадам: $x_i = f(W \cdot v_{ii+k-1} + b)$ қийматлари генерация қилинади, $W \in \mathbb{R}^{d \times k}$. Филтр ойнаси тўлиқ ўтказилганда $m = [x_1, x_2, \dots, x_{n-h+1}]$ ҳосил бўлади. l марта филтрлаш орқали $M = \{m_1; m_2; \dots, m_l\}$ ҳосил қилинади, $M \in \mathbb{R}^{l \times (n-k+1)}$, $E = n - k + 1$.

5-қадам: Ҳужжатдаги сўзларнинг тўлиқ боғлиқлигини қамраб олиш мақсадида M чиқиш векторлари LSTM қатламга узатилади. LSTM яширин қатлами H билан белгиланади. t -қадамдаги учта эшик яъни, кириш i_t , чиқиш o_t ва унутиш f_t қуйидагича янгиланади.

$$\begin{aligned} i_t &= \sigma(W_i m_t + U_i h_{t-1} + b_i), \\ o_t &= \sigma(W_o m_t + U_o h_{t-1} + b_o), \\ f_t &= \sigma(W_f m_t + U_f h_{t-1} + b_f), \\ c_t &= f_t \otimes c_{t-1} + i_t \otimes \tanh(W_c m_t + U_c h_{t-1} + b_c), \\ h_t &= o_t \otimes \tanh(c_t), \end{aligned}$$

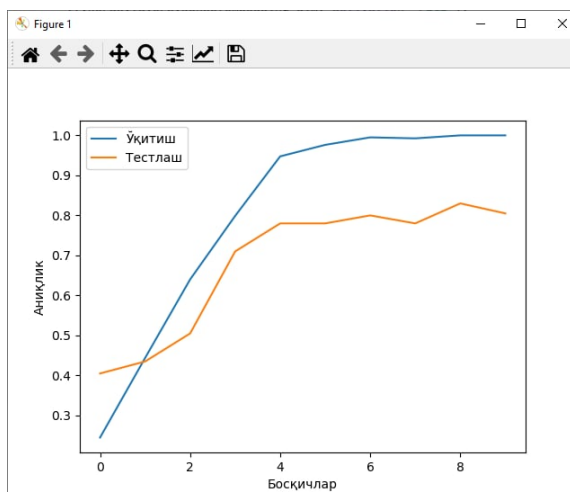
ҳисоблашлар икки йўналишда амалга оширилади. σ бу сигмасимон функция. $W \in \mathbb{R}^{H \times E}$, $U \in \mathbb{R}^{H \times H}$, $b \in \mathbb{R}^H$ тармоқ параметрлари.

6-қадам: H ва H' конкатинацияланиб чиқиш қатламга киритилади.

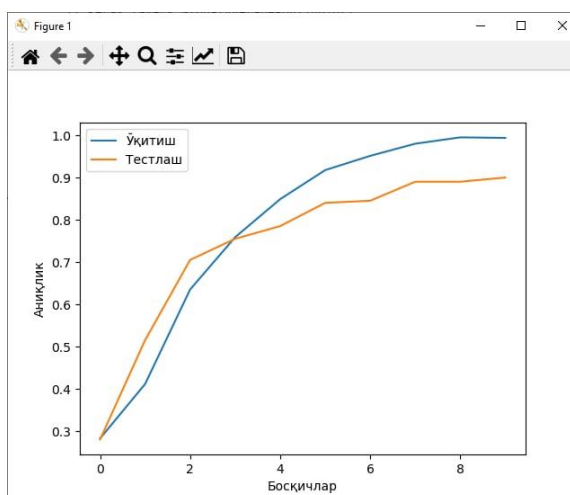
7-қadam: Софтмакс функцияси орқали ҳужжат синфи аниқланади,

$$P_i(y) = \frac{\exp(y_i)}{\sum_{j=1}^K \exp(y_j)}, i=1,2,\dots,K.$$

Олиб борилган тадқиқотлар натижасида чуқур ўқитишнинг такрорий нейрон тармоқ ва конволюцион нейрон тармоқ усуллари мос равишда 80% ва 90% аниқлик билан ишлаши аниқланди (1 ва 2-расмлар).

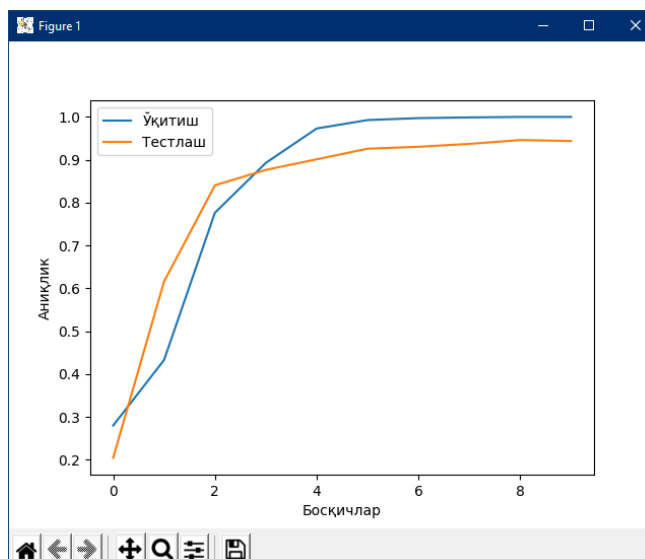


1-расм. RNN усулида олинган натижалар



2-расм. CNN усулида олинган натижалар

Таклиф этилган конволюцион нейрон тармоқ ва такрорий нейрон тармоқ (BLSTM) (гибрид) алгоритми 93% аниқликда ишлаши тажрибавий тадқиқотларда аниқланди (3-расм).



3-расм. Конволюцион нейрон тармоқ ватақрорий нейрон тармоқ (BLSTM) (гибрид) алгоритмида олинган натижалар

Фойдаланилган адабиётлар:

1. О. Ж. Бабомурадов, Л. Б. Бобоев, Х. Т. Дусанов. Матннинг кетма-кетлик модели // *Инновацион ёндашулар илм-фан тараққиёти калити сифатида: ечимлар ва истиқболлар*, ЎЗМУ Жиззах филиали, Республика миқёсидаги илмий-техник анжумани, 2020, pp. 192–197.
2. G. E. Hinton, J. L. McClelland, D. E. Rumelhart. Distributed representations. In the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, Volume 1. Foundations (pp. 77-109).” Cambridge, MA: MIT Press, 1986.
3. D. E. Rumelhart, G. E. Hinton, R. J. Williams. Learning representations by back-propagating errors // *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
4. J. L. Elman. Finding structure in time // *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.
5. T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient estimation of word representations in vector space // *arXiv preprint arXiv:1301.3781*, pp. 1–12, 2013.
6. T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean. Distributed representations of words and phrases and their compositionality // *Advances in neural information processing systems*, 2013, pp. 3111–3119.
7. O. J. Babomuradov, L. B. Boboev, N. N. Abduraxmanova. Hybrid algorithm for classifying text documents // *Олий таълим тизимида масофали таълимни жорий этишнинг техник-дастурий ва услубий таъминотини такомиллаштириш истиқболлари*, Республика илмий-амалий конференцияси, Қариш, 28 май 2021, 2021, pp. 177–179.

ПРОЕКТИРОВАНИЕ МОРФОЛОГИЧЕСКОГО АНАЛИЗАТОРА ДЛЯ ЕСТЕСТВЕННЫХ ЯЗЫКОВ НА ПРИМЕРЕ УЗБЕКСКОГО ЯЗЫКА

Равшанов Н., Бакаев И. И.

Научно-исследовательский институт развития цифровых
технологий и искусственного интеллекта
Бухарский государственный университет

Abstract. A brief review of existing morphological analyzers for agglutinative languages is given. An architecture of a morphological analyzer for the Uzbek language is proposed, an object-oriented model and a diagram of which are described using UML.

Keywords. Computational linguistics, UML, tokenization, stemming, morphonology, state machine.

Аннотация. Приведен краткий обзор существующих морфологических анализаторов для агглютинативных языков. Предложен архитектура морфологического анализатора для узбекского языка, объектно-ориентированная модель и диаграмма вариант использования которого описана при помощи UML.

Ключевые слова. Компьютерная лингвистика, UML, токенизация, стемминг, конечный автомат.

Введение.

Морфологический анализатор является неотъемлемой частью решения подзадач поисковых систем (машинного перевода, информационного поиска, извлечения информации), электронных библиотек, архивов, систем «вопрос-ответ» и документооборота. Морфологический анализатор – это программное обеспечение, которое решает ряд комплексных задач, таких как определение грамматических признаков, основ, лемм, морфем и части речи словоформы рис.1.



Рис.1. Формальная схема морфологического анализатора

Для создания морфологического анализатора существует множество подходов: конечный автомат, конечный автомат с выходом, детерминированный ациклический конечный автомат, подходы основанные на правилах, корпусах, шаблонах, двухуровневой

морфологии и гибридных подходов. На основе их создано множество готовых решений для групп флективных и агглютинативных языков мира Рис.2.

Только определенное количество этих естественных языков обеспечено необходимыми лингвистическими ресурсами и инструментами морфологического анализатора для обработки естественного языка.

Большинство научных исследований основано на комбинаторной смеси нескольких морфологических подходов. Так как в статье будем проектировать морфологический анализатор на примере узбекского языка, поэтому ниже описываются некоторые исследования, относящиеся к агглютинативным языкам.

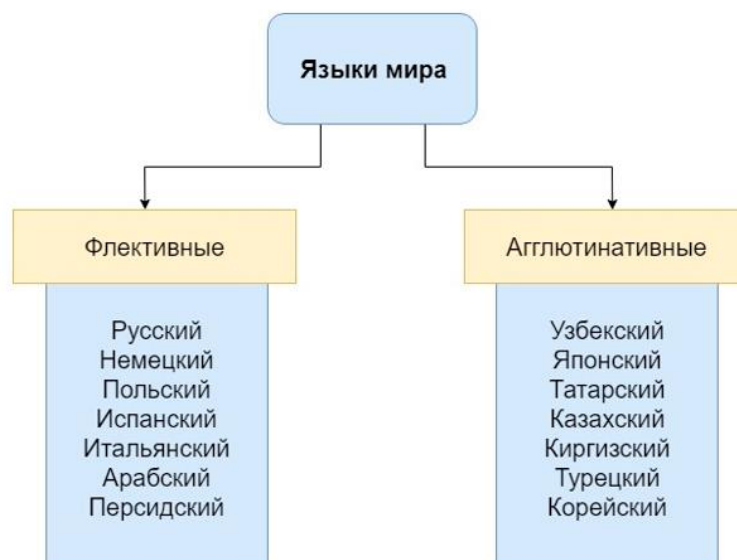


Рис.2. Перечень флективных и агглютинативных языков мира

Связанные работы

Выполняется ряд научных работ по таким языкам, как турецкий, казахский, туркменский, узбекский, киргизский, татарский, относящимся к семье тюркских языков, принадлежащим к группе агглютинативных языков.

В научных работах по турецкому языку были разработаны такие морфологические анализаторы, как TRmorph[1] и TRMOR [2], основанные на подходе двухуровневой морфологии, которые реализованы с помощью SFST. TRmorph состоит трёх компонентов: конечного автомата, представленного регулярными выражениями, набора правил фонологии и орфографии, а также словаря из 37101 слов, принадлежащих 9-ти категориям частей речи. Лексикон морфологического анализатора TRMOR составлен из ресурса Википедии. Когда TRMOR и TRmorph были протестированы с использованием 108 слов, TRMOR дал точность 72%, а TRmorph дал точность 38%[2].

Также авторами [3] описан морфологический анализатор двухуровневой морфологии для турецкого языка, который состоит из пяти модулей: конечный автомат с выходом, механизм правил для суффиксов (файл в формате .xml), лексикон (база данных), структура данных виде дерева и кэш типа LRU (least recently used). Лексикон состоит 54 000 слов, из них 19 000 относится к именам существительных.

В исследованиях [4], [5] разработаны морфологические анализаторы на основе правил для казахского языка. Для реализации правил чередования и морфотактики двухуровневый морфологии используется конечный автомат с выходом Foma и XFST (Xerox Finite State Tools). Основная функция морфологического анализатора устранение неоднозначности для казахского языка, который является разновидностью теггера Брилла.

Также для туркменского языка [6] разработан морфологический анализатор с подходом двухуровневой морфологии. Грамматические правила языка реализованы с помощью конечного автомата типа XFST.

Авторами изложено [7] проектирование и разработка комплекса программных средств который включает лингвистические ресурсы (тезаурусы и онтологии) и инструменты обработки текстов (генератор словоформ, морфологический анализатор и средство морфологического устранения неоднозначности слов) на основе корпусного подхода для казахского языка.

В статье [8] описывается разработанный морфологический анализатор для казахского, татарского и кумыкского языков. Для реализации морфотактических правил используется конечный автомат типа HFST (Helsinki Finite-State Toolkit). Общее количество слов в лексиконе для казахского языка составляет 11224, для татарского – 10737, для кумыкского 4845 слов.

В работе [9] обсуждается инструмент морфологического анализатора для узбекского языка под названием MorphUz. MorphUz работа на основе подхода двухуровневой морфологии, данный подход может определить основу слов [10] и анализировать суффиксы.

В исследовании [11] разработан инструмент морфологического анализа UZMORPP для узбекского языка. Инструмент включает морфотактические и морфофонемные правила узбекского языка, который реализован с помощью логического программирования на языке программирования Prolog.

В работе [12] предложен электронный словарь окончаний узбекских слов для определения морфологических основ слов. Электронный словарь проанализирован с помощью казахского морфологического анализатора типа стемминг [13], [14].

Хотя в представленных выше работах были получены важные результаты фундаментального и практического характера, они не были спроектированы как полноценная система морфологического анализа. Поэтому цель этой задачи спроектировать системы для всех задач морфологического анализа узбекского языка.

Теоретический анализ

Словарный состав узбекского языка постоянно пополняется новыми словами. Пополнение происходит путём образования новых слов с помощью аффиксации и композиции. Аффиксация и композиция являются основными способами образования слов. Аффиксальный способ образования слов наиболее продуктивный в узбекском языке. В этом способе новые слова создаются путём присоединения к корню слова образующих аффиксов. Композиционный способ — это создание новых слов путем сложения двух и более слов. В результате двух указанных выше способов в узбекском языке по структуре и формированию образуются следующие типы слов Рис.3.



Рис.3. Типы слов в узбекском языке по их структуре.

В современном узбекском языке есть шесть самостоятельных частей речи: Имя существительное, Имя прилагательное, Имя числительное, местоимение, глагол, наречие. Каждая часть речи по структуре имеет вышеуказанные типы слов.

На основании вышеприведенной теоретической информации мы определили следующие сущности:

- слова;
- типы слов;
- части речи;
- аффикс;
- типы аффиксов.

Далее определяем сущности для задач морфологического анализатора. К ним относятся следующие:

- Разбиение текста на токены;
- Определение основы слов;
- Определение морфем и грамматических признаков.

Постановка задачи

С учетом проведенного анализа в статье решаются следующие задачи:

- Разработка модели классов морфологического анализатора;
- Построение диаграммы варианта использования для определённых групп пользователей;

Метод решение

Исходя из выше приведенных теоретических материал определяем следующие классы
рис.4:

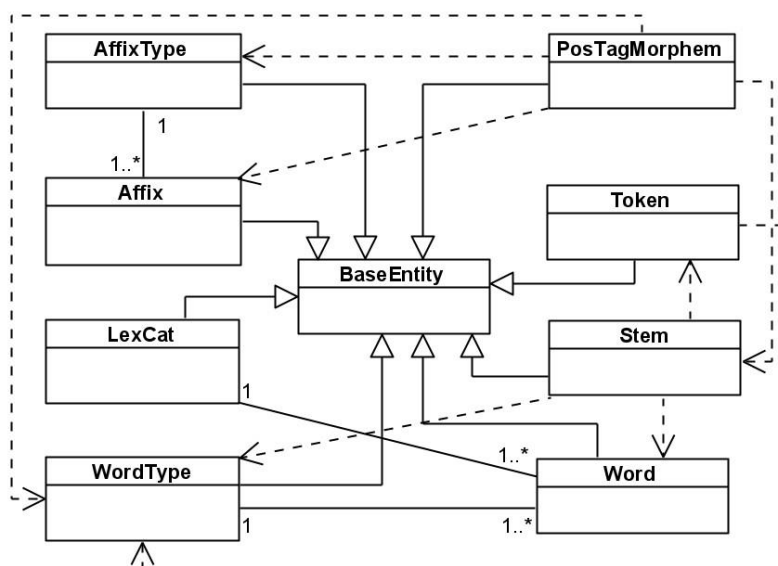


Рис.4. Объектно-ориентированная модель морфологического анализатора

«BaseEntity» — это абстрактный класс, все классы наследуются от этого класса. Этот класс состоит из полей, представляющих идентификатор(id) и значение(value).

«Words» — это класс, представляющий сущность «Слова». У класса есть свои поля «WordParent», чтобы определить, из какого слова оно было создано.

«WordType» — это класс, представляющий сущность «Типы слов».

«Affix» — это класс, представляющий сущность «Аффиксы».

«AffixType» — это класс, представляющий сущность «Типы аффиксов».

«LexCat» — это класс, представляющий сущность «Лексические категории» то есть части речи.

Все классы представленных в диаграммы классов наследоваться от класса BaseEntity. Для «WordType», «Affix», «AffixType», «LexCat» классов достаточно полей, унаследованных от класса BaseEntity. Поэтому у этих классов нет собственных полей. Эти поля могут использоваться для хранения названия типа слов, аффиксов, типа аффиксов, лексических категорий и их идентификаторов.

«Word_LexCat» — это класс, представляющий взаимосвязь между классами «Words» и «LexCat». Связь между классами «Words» и «LexCat» использует дополнительный класс «Word_LexCat», чтобы избежать омонимии некоторых слов.

Связь между классами Affix и LexCat использует класс Affix_LexCat для определения омонимичных суффиксов.

«Stem» — это класс, представляющий сущность «основы слов». Класс состоит из следующих полей и методов:

- "OrfoTypeWord" - метод определения типа слова;
- «StemmingWordW2» - метод, находящий основу слова (основу) типа w_{sb} ;
- "StemmingWordW3W4" - метод нахождения основы (основы) слов типа w_j и w_t ;
- метод "StemmingWordW5", находящий основу слова (основу) типа w_{ab} ;

- метод "StemmingWord", находящий основу (основу) слов типа w_{st} , w_{sy} и w_{psy} ;
- «Lemmatization» - метод, который преобразует словоформы в лемму типа w_{st} , w_{sy} и

w_{psy} ;

- "isprefix" - метод, определяющий, есть ли у слова префикс или нет;
- "prefixLen" - метод определения длины префикса;
- "ReadRule" – метод чтения правил;

Алгоритмы StemmingWordW2, StemmingWordW3W4, StemmingWordW5 и StemmingWord методов реализован в работе [15] ранее.

«Token» — это класс, который разбивает текст на слова. Класс содержит метод «Tokenlash»[16], который разбивает текст на слова.

«POSTagMorphem» — это класс, который определяет слово в морфемы и свойства морфем[17]. Класс состоит из следующих методов:

- «isbelogswordlang» — метод, проверяющий принадлежность слова к языку;
- "POSTaggingMorphemic" - метод маркировки аффиксов грамматическими тегами;
- "MorphemicList" - метод, формирующий список аффиксов;
- "MorphemicTypeList" - метод, формирующий тип аффиксов;

Класс «POSTagMorphem» использует отношение «Dependency» (зависимость) с классами «AffixType», «Affix» и «WordType», то есть в качестве входных и выходных значений.

Далее построим диаграмму вариант использования для определённых групп пользователей.

Для системы морфологического анализатора можно выделить следующие группы пользователей:

- пользователи (студент, преподаватель, API пользователь) – может провести морфологический анализ (задачи токенизация, определение основ слов, морфем и грамматических признаков) слов с помощью системы рис.5а.

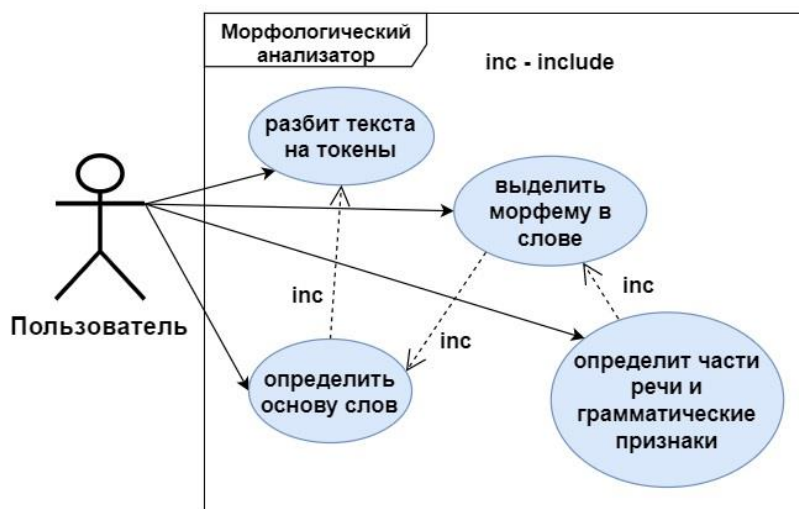


Рис.5а. Диаграмма вариант использования морфологического анализатора для пользователей

- эксперт-лингвист – специалист в области компьютерный лингвистики который может сделать операция CRUD (аббревиатура англ. слов create, read, update, delete) рис.5b.

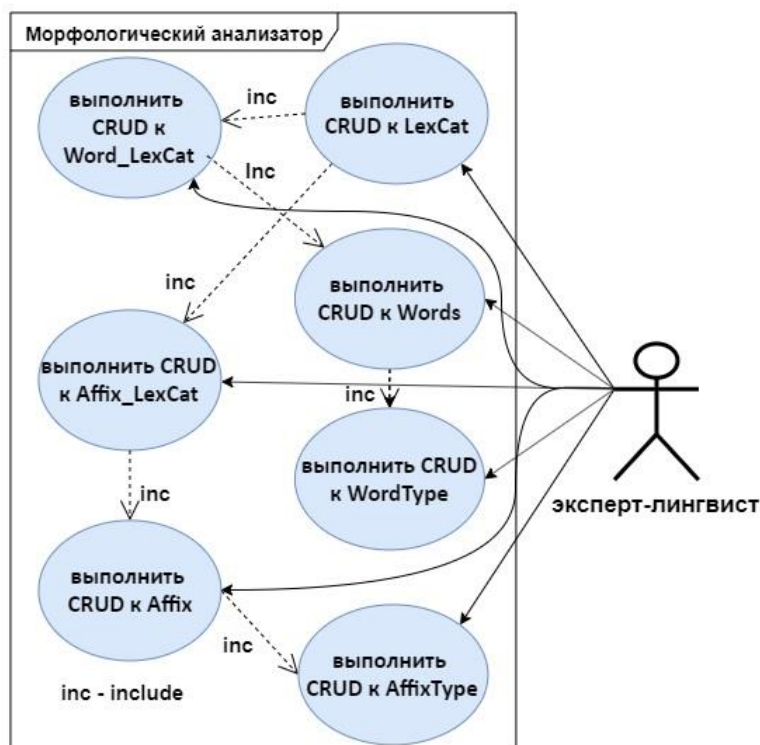


Рис.5b. Диаграмма вариант использования морфологического анализатора для эксперт-лингвистов.

Результаты

На основе применения разработанного морфологического анализа слов узбекского языка для текстовых запросов на узбекском языке при поиске информационных ресурсов в системе электронной библиотеки Бухарского областного информационно-библиотечного центра имени Абу Али ибн Сино и в центре информационных ресурсов Бухарского государственного университета обеспечена возможность сокращения времени и трудозатрат на 9-11% при выполнении операций составления библиографических описаний документов на узбекском языке в системе электронного каталога за счет автоматической коррекции орфографических ошибок и выдачи правильного варианта написания слов.

Заключение

По итогам проведённого анализа научных исследований, изложенных в работе, связанных с созданием морфологических анализатора для узбекского языка, относящихся к семье агглютинативных языков, разработаны модели классов морфологического анализатора, и построена диаграмма варианта использования морфологического анализатора для определённых групп пользователей.

Использованная литература:

1. G. Gagri, "A Freely Available Morphological Analyzer for Turkish," in *LREC*, 2010, pp. 820–827. Accessed: Jun. 29, 2020. [Online]. Available: <http://www.gnu.org/licenses/gpl.html>
2. K. Ayla, S. Helmut, T. Ahmet E, and K. Özkan, "TRMOR: a finite-state-based morphological analyzer for Turkish," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 27, pp. 3837–3851, 2019, doi: 10.3906/elk-1902-125.
3. O. T. Yıldız, A. Begüm, and E. Gökhan, "An Open, Extendible, and Fast Turkish Morphological Analyzer," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 2019, pp. 1364–1372.
4. G. Kessikbayeva and I. Cicekli, "A Rule Based Morphological Analyzer and A Morphological Disambiguator for Kazakh Language," *Linguist. Lit. Stud.*, vol. 4, no. 1, pp. 96–104, 2016.
5. G. Kessikbayeva and I. Cicekli, "Rule Based Morphological Analyzer of Kazakh Language," in *Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM*, 2014, vol. Proceeding, pp. 46–54.
6. A. C. Tantug, E. Adalı, and K. Oflazer, "Computer Analysis of the Turkmen Language Morphology," in *Advances in Natural Language Processing*, 2006, pp. 186–193.
7. D. Akhmed-Zaki, M. Mansurova, G. Madiyeva, N. Kadyrbek, and M. Kyrgyzbayeva, "Development of the information system for the Kazakh language preprocessing," *Cogent Eng.*, vol. 8, no. 1, p. 1896418, 2021, doi: 10.1080/23311916.2021.1896418.
8. J. Washington, I. Salimzyanov, and F. Tyers, "Finite-state morphological transducers for three Kypchak languages," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation ({LREC}'14)*, 2014, pp. 3378–3385. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2014/pdf/1207_Paper.pdf
9. N. Abdurakhmonova, I. Alisher, and R. Sayfulleyeva, "MorphUz: Morphological Analyzer for the Uzbek Language," in *2022 7th International Conference on Computer Science and Engineering (UBMK)*, 2022, pp. 61–66. doi: 10.1109/UBMK55850.2022.9919579.
10. D. Mengliev, V. Barakhnin, and N. Abdurakhmonova, "Development of Intellectual Web System for Morph Analyzing of Uzbek Words," *Appl. Sci.*, vol. 11, no. 19, 2021, doi: 10.3390/app11199117.
11. G. Matlatipov and V. Zygmunt, "Representation of Uzbek Morphology in Prolog," *Asp. Nat. Lang. Process.*, vol. 5070, pp. 83–110, 2009, doi: 10.1007/978-3-642-04735-0.
12. S. Matlatipov, U. Tukeyev, and M. Aripov, "Towards the Uzbek Language Endings as a Language Resource," in *Advances in Computational Collective Intelligence*, 2020, pp. 729–740.
13. U. Tukeyev, A. Turganbayeva, B. Abduali, D. Rakhimova, D. Amirova, and A. Karibayeva, "Lexicon-free stemming for Kazakh language information retrieval," in *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*, 2018, pp. 1–4. doi: 10.1109/ICAICT.2018.8747021.
14. U. Tukeyev, A. Karibayeva, A. Turganbayeva, and D. Amirova, "Universal Programs for Stemming, Segmentation, Morphological Analysis of Turkic Words," in *Computational Collective Intelligence*, 2021, pp. 643–654.

15. I. . Bakayev, “Development of a stemming algorithm based on a linguistic approach for words of the uzbek language,” in *International Conference on Scientific, Educational & Humanitarian Advancements*, 2021, pp. 195–202.
16. I. I. Bakaev, “Linguistic features tokenization of text corpora of the Uzbek language,” *Bull. TUIT Manag. Commun. Technol.*, vol. 4, 2021
17. Н. Равшанов, И. . Бакаев, and Т. . Шафиев, “Особенности разработки морфологического анализатора узбекского языка,” *Проблемы вычислительной и прикладной математики*, vol. 4, no. 28, pp. 121–131, 2020.

ANALYSIS OF THE SEMANTIC RELATEDNESS OF WORDS BASED ON A
COMMON DICTIONARY

Tuliyev Ulug'bek

Senior teacher

National University of Uzbekistan

Annotation: A technique for creating sets (groups) of natural language words (concepts) based on data from textual documents is considered. It is believed that the set of documents is divided into two disjoint classes, which are described using a common dictionary. Sets of words are used to reduce the dimension of the initial space through the formation of latent features on them. The choice of latent features is justified by the values of the content authenticity of documents. Based on the ranking of features based on the results of splitting documents into clusters, semantically related words are selected by topic.

Keywords: Thematic modeling, latent feature, ontology, common dictionary

Аннотация: Рассматривается методика создания наборов (групп) из слов (понятий) естественного языка по данным из текстовых документов. Считается, что множество документов разделено на два непересекающихся класса, для описания которых используется общий словарь. Наборы слов применяются для снижения размерности исходного пространства через формирование по ним латентных признаков. Выбор латентных признаков обосновывается по значениям контентной аутентичности документов. На основе ранжирования признаков по результатам разбиения документов на кластеры производится отбор семантически связанных слов по темам.

Ключевые слова: Тематическое моделирование, латентные признаки, онтология, общий словарь

Introduction.

Machine learning methods based on vector models are successfully used to extract knowledge from text documents in various subject areas. With a vector representation in the form of frequency distributions of words in contexts, it is considered that semantically close words have close vectors. As an option for testing models, manual formation of sets of pairs of words with expert estimates of proximity is considered.

In probabilistic thematic modeling [2], the possibilities for experts to interpret the content of topics are not limited by the values of the semantic similarity of words. Optimization of splitting documents into topics is achieved through the use of additional criteria-regularizers. To interpret the partitioning results, a list of the most likely words specific to each topic is analyzed. As a rule, the contents of the list from subject topics and the almost complete disregard of lists from background topics are of practical interest.

The interpretation of the meaning of words from dictionaries in different subject areas requires the involvement of specialized resources for analysis - ontologies and the introduction of certain restrictions. Building ontologies of the corresponding subject areas [4] is a common way to formalize different areas of knowledge.

The distribution of data in topic modeling problems is generally considered to be unknown. The presence of outliers can greatly affect the results of cluster analysis. In order to reduce the impact of outliers, it is proposed to use the achievement of robust stability through data grouping based on the results of nonlinear feature transformations [5].

One of the goals of this study is to develop new topic modeling methods for analyzing the semantic relatedness of words when presenting text data through a bag of words model [1]. It is not known in advance how many topics to set the model, since this number is a free parameter [3]. Since the tasks being solved are incorrect, the number of topics and their composition is determined on the basis of additional conditions. Conditions establish: occurrence of words in the general dictionary; the number of words in the general dictionary; breaking down documents into topics.

Description of documents through latent features allows you to:

- reduce the dimension of the feature space;
- to implement training of a thematic model on a common dictionary of collections from two classes of documents;
- use traditional metric algorithms for cluster analysis of documents;
- justify the choice of the number of topics based on the results of cluster analysis;
- identify sets of semantically related words by topic.

Similarly [3], a collection of documents for machine learning, when divided into two non-overlapping classes, can be represented as:

- the union of two non-overlapping subsets, each of which contains documents from one or more subject areas;
- sets of articles from two periodicals, each of which represents a separate class.

Statement of Problem.

Given a sample of objects $E_0 = \{S_1, \dots, S_m\}$, divided into 2 non-overlapping classes K_1, K_2 . Objects in E_0 are described by a set of quantitative features $X(n) = (x_1, \dots, x_n)$, whose values represent the frequency of occurrence of words in documents.

Procedures are defined for:

- calculation of a set of latent features $Y(t) = (y_1, \dots, y_t)$ over non-overlapping subsets $X(n_1), \dots, X(n_t)$, $t \geq 1$, $\forall c \in \{1, \dots, t\}$, $n_c > 1$, $n_1 + \dots + n_t \leq n$;
- partitioning of objects E_0 into a given number of non-overlapping clusters (groups) $G_1, \dots, G_p$.

Required:

- evaluate the quality of partitioning $F(p, Y(\sigma))$ into p groups according to the set of latent features $Y(\sigma) \subset Y(t)$;
- select the number of topics for clustering groups $G_1, \dots, G_p$;
- identify sequences of words for each topic.

To solve the problem, clustering of raw features (frequency representations of words) and objects (documents) is used. The process of forming a latent feature space and selecting a set of words for a common vocabulary for two classes of documents is designed to use several criteria. The formation of sets of latent features is associated with:

- non-linear transformation of raw feature values into a unified representation in $\{1, 2\}$;
- calculation of stability values for each feature;

–creation of non-overlapping sets of raw features according to an ordered sequence of their stability values;

– calculation of latent features for each set.

Nonlinear feature transformations and calculation of generalized estimates of objects.

The fundamental properties of the criteria considered below for dividing feature values into intervals are the invariance of their results to the measurement scales and the low variability of stability on different samples from the general population [6]. The results of splitting into intervals are used to convert feature values to gradations in the nominal scale and calculate the measure of compactness by latent features.

The conditions for combining words into a common dictionary and the formation of a set of latent features for splitting documents into groups are determined. It is proposed to choose the number of document topics in relation to the connectivity of group centers.

The essence of non-linear transformations of attributes is reduced to replacing their initial values with the values of the membership function of objects to classes. In order to unify the notation, instead of $d_{tc}(u,v)$, $t=1,2$ for the interval $[r_u; r_v]^\mu$ in $x_c \in X(n)$, we will use $d_{tc}(\mu)$. The value of the membership function $f_c(\mu)$ to the class K_1 over the interval $[r_u; r_v]^\mu$ (gradations $\mu \in \{1, \dots, p_c\}$) is calculated as

$$f_c(\mu) = \frac{d_{1c}(\mu)/|K_1|}{d_{1c}(\mu)/|K_1| + d_{2c}(\mu)/|K_2|}. \quad (1)$$

Replacing the feature gradations with the values of the membership function of objects to the class K_1 according to (1) for $p_c > 2$ is considered as a non-linear transformation, in which the sequence order on the numerical axis is determined. When calculating the membership function $f_c(\mu)$ to the class K_1 according to the gradation $\mu \in \{1, 2, \dots, p_c\}$, the number of objects of the class K_1 (K_2) with the value μ is used as $d_{1c}(\mu)$ ($d_{2c}(\mu)$). We will denote by D the set of feature indices from the set $X(n)$ that have nonzero values.

Computational experiment.

The computational experiment was carried out on textual data from 12 subject areas, represented by many abstracts of the Higher Attestation Commission of the Republic of Uzbekistan [9, 10, 11]. For the description, 473 words were used from a combination of 12 dictionaries in subject areas, compiled by an expert. The demand for subject dictionaries was in their use for testing the results of the experiment in the formation of general dictionaries for collections of documents.

Sequences of words on the topic from class K_1 (Chemistry_Technology)

№	№	№	№
1	Нуклеофил	9	Электрод
2	Юридик	10	Адаптация
3	Бадий	11	Гибрид
4	Фосфор	12	дигидрохиназолин
5	Потенциал	13	Препарат
6	Матрица	14	Реабилитация
7	Фрагмент	15	Магнит
8	Аналитик	16	Гипертензия

Conclusion.

A technology for the formation of general dictionaries and sets of semantically related sequences of words by topic is proposed. The technology is based on the use of splitting collections of documents into two non-overlapping classes. To implement the technology, grouping of words was used in order to form a latent feature space and group documents in this space.

References:

1. 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), Tashkent, Uzbekistan, 2021, pp. 1-4, doi: 10.1109/ICISCT52966.2021.9670043.
2. Vorontsov K. V., Potapenko A. A. Additive Regularization of Topic Models // Machine Learning Journal. — 2015. — Vol. 101. — Pp. 303–323.
3. Краснов Ф. В., Диментов А. В., Шварцман М. Е. Использование тематических моделей для парного сравнения коллекций научных статей *Информ. и её примен.*, **14:3** (2020), 129–135.
4. Дорофеюк А.А., Покровская И.В., Чернявский А.Л. Структуризация объектов нечисловой природы // Информационные технологии и вычислительные системы 1/2018. С.16 –21.
5. Ignatiev N. A. On Nonlinear Transformations of Features Based on the Functions of Objects Belonging to Classes // Pattern Recognition and Image Analysis. 2021. V. 31. № 2. P. 197–204.
6. Петровский А.Б., Лобанов В.Н. Многокритериальный выбор в пространстве признаков большой размерности: мультиметодная технология ПАКС-М // Искусственный интеллект и принятие решений 3/2014. С. 92–104.
7. Игнатъев Н. А., Рахимова М. А. Формирование и анализ наборов информативных признаков объектов по парам классов // Искусственный интеллект и принятие решений 4/2021. С. 18–26.
8. Ignatyev N. A. Structure Choice for Relations between Objects in Metric Classification Algorithms // Pattern Recognition and Image Analysis. 2018. V. 28. № 4. P. 590–597.
9. Ignatev N.A., Tuliev U.Y. Semantic structuring of text documents based on patterns of natural language entities // Computer Research and Modeling, 2022, vol.14, no. 5, pp.1185-1197 DOI: 10.20537/2076-7633-2022-14-5-1185-1197
10. Тулиев У. Ю., Лолаев М. Я. О формировании пространства для описания тематических документов // Серия «Информатика. Информационная безопасность. Математика». 2021. № 1 35-50.

KORPUSDA MORFOLOGIK ANALIZATOR UCHUN LINGVISTIK MODELLASHTIRISH MASALALARI

Urazaliyeva Mavluda Yangiboyevna

O'zbekiston Milliy universiteti,

Kompyuter lingvistikasi yo'nalishi magistranti,

E-mail: mavludaaurazaliyeva@gmail.com

Annotatsiya: Ushbu maqolada bugungi kunda kompyuter texnologiyalari yordamida yaratilayotgan korpuslar, ulardagi imkoniyatlar asosida matnlarni sodda va sifatli tahlil qilish, uzbekcorpus.uz platformasi interfeysi, mavjud qulayliklari, korpus yaratishdagi lingvistik baza va unda so'zlarni morfologik tahlil qilishdagi nazariyalar, modellar, so'z turkumlarining turlicha qoliplanishi, shuningdek xorijiy korpuslardagi lingvistik modellar bilan o'zaro mutanosibligi tahlil qilingan. Morfoanalizator barcha tillarga xos bo'lgan va kompyuter dasturlash tillari uchun qulay bo'ladigan sodda qoliplar yaratilishi asosiy ustunlikka egaligi ta'kidlab o'tilgan.

Kalit so'zlar: elektron korpus, korpusshunoslik, morfoanaliz, lingvistik baza, lingvistik modelashtirish, korpus interfeysi.

Аннотация: В данной статье рассмотрены имеющиеся на данный момент корпуса, простой и качественный анализ текстов с учетом их возможностей, интерфейс платформы uzbekcorpus.uz, ее существующие возможности, лингвистическая база для создания корпуса и теории, модели и морфологические анализ слов в нем, различное формообразование словосочетаний, а также иноязычное соотнесение с лингвистическими моделями в корпусах. Отмечено, что основным преимуществом морфоанализатора является создание простых шаблонов, характерных для всех языков и удобных для языков компьютерного программирования.

Ключевые слова: электронный корпус, исследование корпусов, морфоанализ, лингвистическая база, лингвистическое моделирование, интерфейс корпуса.

Abstract: This article discusses currently available corpora, simple and high-quality analysis of texts, taking into account their capabilities, the interface of the uzbekcorpus.uz platform, its existing capabilities, the linguistic basis for creating a corpus and theory, models and morphological analysis of words in it, various formation of word combinations, as well as foreign language correlation with linguistic models in corpora. It is noted that the main advantage of the morphoanalyzer is the creation of simple patterns that are typical for all languages and convenient for computer programming languages.

Keywords: electron corpus, corpus research, morphoanalysis, linguistic base, linguistic modeling, corpus interface.

Zamonaviy axborot texnologiyalaridan foydalanish tilning turli imkoniyatlaridan foydalanish yo'lida katta imkoniyatlar eshigini ochdi. Kompyuter tarjimas, matnlarning tahlili, elektron lug'atlar, elektron korpuslarni bunga misol qilib keltirish mumkin. Elektron lug'atlar

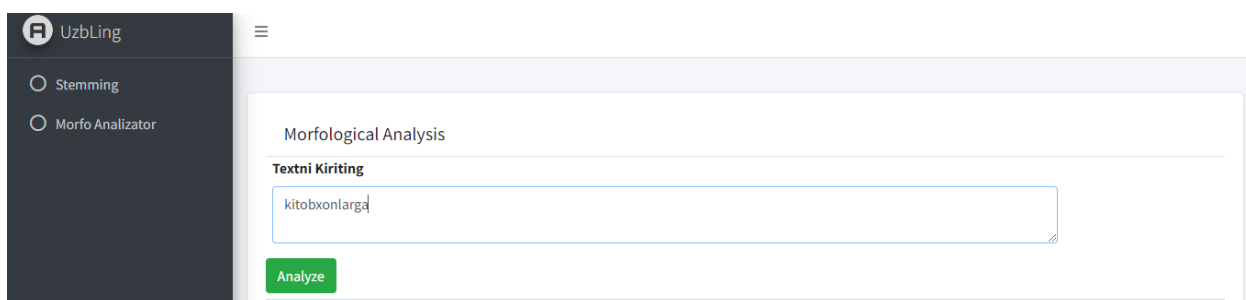
tuzish va ularni amaliyotda qo'llash tilshunoslar va shu sohaga doir barcha shaxslarga yengillik berdi va undan foydalanish ko'nikmasini yaratdi.

Bugungi kunga kelib til texnologiyasida erishilgan ilmiy natijalardan unumli foydalanish, ularga yangicha yondashuvlar, kompyuter texnologiyalaridan unumli foydalanish ilmiy muammolarni oson va tez hal qilishga yordam beradi. Har xil soha vakillarini bir joyda to'plashda asosiy vosita bo'lib ham xizmat qiladi.

Tabiiy tillar haqida gap borar ekan bevosita uni tashkil qiluvchi birliklarga e'tibor qaratamiz: lisoniy modellar va nutqiy birliklar. Til doimo o'zgarishlarga uchraydi va rivojlanib boradi, ijtimoiylashadi. Barcha tillar o'ziga xos morfologik tuzilishga ega. So'zlarni yasash, ularni matnlarda qo'llash inson ongiga bog'liq. Har bir til esa minglab so'zlarni qamrab oladi. Mana shu so'zlar orqali muloqotga kirishganda turli nutqiy hosilalardan so'z shakllari yuzaga chiqadi. "Kompyuter morfologiyasi ana shu nutq aktidagi birliklarni jarayondagi mexanizmlarni aniqlash" va tahlil qilish bilan shug'ullanadi.

Morfologiyaning eng kichik birligi morfemadir. Morfologiyada so'zlarning eng kichik ma'noli qismigacha bo'lgan birlik aniqlanadi. Turli tillarda ham bu holatlar turlicha. Tabiiyki, o'zbek tilida yaratilayotgan korpus uchun ham morfem jihatdan tahlil eng kerakli qismlaridan biri. Kompyuter tiliga mos ravishda so'zlarni morfem tahlil qilib berish matnlarni to'g'ri tarjima qilishdagi birinchi qadamlardan biridir. Jahonda morfologik sintez bo'yicha ko'pgina olimlar tomonidan ilmiy izlanishlar qilingan. Respublikamizda o'zbek tilidagi matnni morfologik tahlil qiluvchi dasturning lingvistik ta'minotini yaratishga qaratilgan bir qancha tadqiqotlar olib borilgan bo'lib, A.Po'latov, N.Jo'rayeva, B.Alixojayevlarning o'zbek tilini modellashtirishga oid "Разработка формальной модели грамматики узбекского языка" va "Разработка программы компьютерного анализа и синтеза глаголов узбекского языка" nomli tadqiqotlarini, qolaversa N.Abdurahmonovning "Mashina tarjimasining lingvistik ta'minoti" nomli monografiyasi va "Kompyuter lingvistikasi" darsligida morfologik tahlil uchun modellashtirish, formal til nazariyasi, o'zbek tilining grammatik lingvistik bazani tuzishdagi muhim jihatlari ochib berilganligini, Sh.Xamroyevaning "O'zbek tili morfologik analizatorining lingvistik ta'minoti" dissertatsiyasida turkiy tillarda ishlab chiqilgan morfologik analizatorlar funksional imkoniyatining lingvistik ta'minot jihatidan farqli, tilga avtomatik ishlov berishning morfologik tahlil usullari jihatidan esa o'xshash ekanligi aniqlanganini ko'rish mumkin. [Xamrayeva.Sh – 2021:45b]

O'zbek tilining korpusi bo'lgan uzbekcorpus.uz platformasida mavjud morfologik analiz quyida ko'rsatilgan bo'lib, ushbu dastur so'zni morfologik tahlil qilishda chapdan o'ngga qarab analiz qilish algoritmiga asoslangan:



morfologiya

Input: kitobxonlarga->Root word: kitobxon->Turkum: NOUN->Suffix: larga->Analiz noun: lar+ga Modeli: ->Plur+Dat

Showing 1 to 1 of 1 entries

Previous 1 Next

Morfologik tahlil qilish sintaktik analiz uchun ham juda muhim. Bu esa mashina tarjimasining tez va sifatli bo'lishida ham muhim ahamiyat kasb etadi. Bunda so'zning qaysi turkumga oidligini aniqlab olish zarur. Kompyuterlashtirilgan morfologiya asosiy vazifasi "kiruvchi belgilar yoki fonemalardan iborat so'zlarni analiz holatida chiquvchiga yetkazish". [Abduraxmonova N.Z. – 2021:] So'zlarni morfemalarga ajratish quyidagicha: **Ot: bola+jon+lar+im+dagi+lar+niki+mas+mi+kan+a;** **sodda** **fe'l: o'qi+t+tir+ma+yot+gan+lig+i+ni;** **qo'shma fe'l: mashq qil+dir+ish+ayot+gan+lar;** **ko'makchi fe'lli so'z qo'shilmasi: ber+dir+tir+ib yub+or+ma+yot+gan+dan+mi+kan+a** va boshqalar.

Inson uchun so'zlarning morfologik shakllanishi, ya'ni uni nutqda voqeanishi, so'z asosi va qo'shimchalarga ajratish tabiiy hodisa. Biroq so'zlarning avtomatik tahlilida bu protsedura sifatida qaraladi. [Abduraxmonova N.Z – 2021:160b]

morfologiya

Input: mashq->Root word: mashq->Turkum: NOUN->Suffix: NONE->Analiz Noun: NONE

Input: qildirshayotganlar->Root word: qil->Turkum: NOUN->Suffix: dirshayotganlar->Analiz noun: dir Modeli: ->PM+PART

Input: qildirshayotganlar->Root word: qildir->Turkum: VERB->Suffix: shayotganlar->Analiz Verb: sh+ayotgan+lar Modeli: ->Rcp+Vnoun+Rfl+Part+Rfl+Pass+FUT+Cau+Part

Showing 1 to 3 of 3 entries

Previous 1 Next

Morfologik analizator har bir birlikni kontekstualogik jihatdan to'g'ri aniqlab berishi kerak. O'zbek tilida fe'lga qo'shilishi mumkin bo'lgan qo'shimchalar quyidagilar:

Inkor shakl-N {-ma| -mas| -may}
Shart mayli-CF {-sa}
Zamon -T {-a|-y|-yap|-moqda|-yotir...}
Kesimlik shakllari -PF {-man|-san|-dir...}
Nisbat -VF {-t|-tir|-giz, |-kiz...}
Fe'ning vazifadosh shakllari -NF {-gan|-kan|-qan|-b|-ib...}
Shaxs-son P {-im|-ing|-k|-ngiz|-lar| -man| -san|-k|-ngiz}
Maqsad -PS {-moqchi} [Abduraxmonova N.Z. – 2021: 162b]

Bir qancha so'zlarni tahlil qilish jarayonlarini quyida ko'rish mumkin:

 UzbLing

☐ Stemming
 ☐ Morfo Analizator

Morphological Analysis

Textni Kiriting

Analyze

morfologiya

Input:aylanmachoq->Root word: aylan->Turkum: VERB->Suffix:machoq->Analiz Verb:ma Modeli: ->Neg

Showing 1 to 1 of 1 entries

Previous 1 Next

Morphological Analysis

Textni Kiriting

Analyze

morfologiya

Input:vatandoshlarimiz->Root word: vatandosh->Turkum: NOUN->Suffix:larimiz->Analiz noun:lar+imiz Modeli: ->Plur+Poss1

Showing 1 to 1 of 1 entries

Previous 1 Next

Analyze

morfologiya

Input:ertalabgacha->Root word: ertalab->Turkum: ADV

Showing 1 to 1 of 1 entries

Previous 1 Next

Yuqorida ko'rsatilgan so'zlar morfologik tahlil qilinishda piton dasturlash tilidan foydalanilgan bo'lib, bunda asosi, uning qaysi turkumga oidligi va qo'shimchalari ajratib ko'rsatiladi. Jumladan:

Input:aylanmachoq->Root word: aylan->Turkum: VERB->Suffix:machoq->Analiz Verb:ma Modeli: ->Neg;

Input:vatandoshlarimiz->Root word: vatandosh->Turkum: NOUN-> Suffix:larimiz-> Analiz noun:lar+imiz Modeli: ->Plur+Poss1;

Input:ertalabgacha->Root word: ertalab->Turkum: ADV ko'rinishida.

Ma'lumotlar bazasida berilgan so'zlar orfografik jihatdan to'g'ri yozilishi ham hisobga olingan bo'lib, morfoanalizator to'g'ri va sifatli ishlashiga zamin yaratilgan. Bundan tashqari platformada algoritm aniq ishlashi uchun qo'shma so'zlar, juft va takroriy so'zlar uchun alohida morfologik qolip ishlab chiqilmoqda.

Foydalanilgan adabiyotlar:

1. N. Abdurakhmonova, I. Alisher and R. Sayfulleyeva, "MorphUz: Morphological Analyzer for the Uzbek Language," 2022 7th International Conference on Computer Science and Engineering (UBMK), 2022, pp. 61-66, doi: 10.1109/UBMK55850.2022.9919579.
2. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
3. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
4. Abdurakhmonova N., Tuliyeu U. and Gatiatullin A., "Linguistic functionality of Uzbek Electron Corpus: uzbekcorpus.uz," 2021 International Conference on Information Science and Communications Technologies (ICISCT), 2021, pp. 1-4, doi: 10.1109/ICISCT52966.2021.9670043.
5. 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.
6. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.
7. Ш.Хамроевнинг "Ўзбек тили морфологик анализаторининг лингвистик таъминоти" диссертацияси. Тошкент,2021.
8. Бакаев И.И. — Разработка алгоритма стемминга для слов узбекского языка // Кибернетика и программирование. – 2021
9. <https://uzbekcorpus.uz/>
10. <https://uzbeklinganalysis.uz/morfologiya>

RECENT ADVANCEMENTS IN UZBEK NLP TOOLS

Shaymardanov Abror

Uzbekistan State World Languages University

Annotation: This article explores recent advancements in Uzbek natural language processing (NLP) AI tools and their potential impact on language processing tasks in Uzbekistan. The article highlights the challenges faced by NLP tools in Uzbek, such as a lack of linguistic resources, limited research and development, and a shortage of qualified experts. However, the article also presents recent breakthroughs in Uzbek NLP AI tools, including the development of large-scale datasets, deep learning models, and automated language translation systems. These advancements have significantly improved the accuracy and efficiency of language processing tasks, making them more accessible and effective for various applications, including sentiment analysis, speech recognition, and machine translation. The article concludes that the recent advancements in Uzbek NLP AI tools offer promising opportunities for improving language processing tasks in Uzbekistan, contributing to the growth of the country's digital economy and advancing its role in the global AI market.

Keywords: NLP, AI, Linguistics, Computational linguistics, Language

Аннотация: В этой статье рассматриваются последние достижения в области инструментов искусственного интеллекта для обработки естественного языка (NLP) на узбекском языке и их потенциальное влияние на задачи обработки языка в Узбекистане. В статье освещаются проблемы, с которыми сталкиваются инструменты НЛП на узбекском языке, такие как нехватка лингвистических ресурсов, ограниченные исследования и разработки, а также нехватка квалифицированных экспертов. Тем не менее, в статье также представлены последние достижения в инструментах искусственного интеллекта узбекского НЛП, включая разработку крупномасштабных наборов данных, моделей глубокого обучения и автоматизированных систем языкового перевода. Эти усовершенствования значительно повысили точность и эффективность задач языковой обработки, сделав их более доступными и эффективными для различных приложений, включая анализ настроений, распознавание речи и машинный перевод. В статье делается вывод о том, что последние достижения в узбекских инструментах NLP AI открывают многообещающие возможности для улучшения задач обработки языка в Узбекистане, способствуя росту цифровой экономики страны и повышению ее роли на мировом рынке AI.

Ключевые слова: НЛП, ИИ, лингвистика, компьютерная лингвистика, язык.

Introduction.

Natural Language Processing (NLP) is a critical area of Artificial Intelligence (AI) that focuses on enabling machines to understand, interpret, and generate human language. NLP has significant potential for various applications, such as sentiment analysis, speech recognition, and machine translation, that can benefit multiple industries and sectors. However, developing NLP AI tools for languages other than English has been challenging, particularly for languages that lack

linguistic resources, research and development, and qualified experts. One such language is Uzbek, spoken by over 30 million people worldwide. Uzbek is a Turkic language that uses the Cyrillic alphabet and has complex morphology and syntax, making it challenging for NLP AI tools to process. Furthermore, Uzbek has a limited digital footprint, with little research and development in the field of NLP. However, in recent years, there have been significant advancements in Uzbek NLP AI tools that have the potential to transform language processing tasks in Uzbekistan. The development of large-scale datasets, deep learning models, and automated language translation systems has significantly improved the accuracy and efficiency of language processing tasks, making them more accessible and effective for various applications. The purpose of this article is to explore recent advancements in Uzbek NLP AI tools and their potential impact on language processing tasks in Uzbekistan. The article will begin by discussing the challenges faced by NLP tools in Uzbek, including a lack of linguistic resources, limited research and development, and a shortage of qualified experts. It will then present recent breakthroughs in Uzbek NLP AI tools, including the development of large-scale datasets, deep learning models, and automated language translation systems. Finally, the article will conclude by discussing the potential opportunities and challenges presented by the recent advancements in Uzbek NLP AI tools.

Discussion.

One of the significant challenges facing NLP tools in Uzbek is a lack of linguistic resources. Unlike languages such as English, which have a vast amount of linguistic resources, including text corpora, annotated datasets, and lexicons, Uzbek has limited digital resources. This lack of resources makes it challenging to develop accurate and efficient NLP AI tools for Uzbek. Furthermore, Uzbek has a complex morphology and syntax, which requires specific linguistic expertise and resources to develop effective NLP AI tools. Another challenge facing Uzbek NLP AI tools is limited research and development in the field. Uzbekistan has a relatively small research community, and there is a shortage of qualified experts in the field of NLP. As a result, there has been limited research and development in NLP AI tools for Uzbek. This lack of research and development has slowed down the progress of Uzbek NLP AI tools and has limited their potential impact.

Despite these challenges, recent advancements in Uzbek NLP AI tools offer promising opportunities for improving language processing tasks in Uzbekistan. One of the significant breakthroughs in Uzbek NLP AI tools is the development of large-scale datasets. These datasets provide a significant amount of annotated text data for training NLP AI models, improving their accuracy and efficiency. The availability of these datasets has significantly improved the quality of NLP AI tools for Uzbek. Another breakthrough in Uzbek NLP AI tools is the development of deep learning models. Deep learning models use artificial neural networks to process and analyze large amounts of data, making them effective for language processing tasks. The development of deep learning models has significantly improved the accuracy of Uzbek NLP AI tools, making them more effective in applications such as sentiment analysis and speech recognition. Finally, the development of automated language translation systems has the potential to transform language processing tasks in Uzbekistan. These systems use machine learning algorithms to translate text from one language to another, making them effective for cross-lingual communication. Automated language translation systems can also be used to translate Uzbek text into other languages, improving the accessibility of Uzbek language content to a global audience. The recent

advancements in Uzbek NLP AI tools have opened up new opportunities for developing innovative applications in Uzbekistan. Two notable examples of these applications are UzbekVoice AI and MohirAI. UzbekVoice AI is an AI-powered voice assistant developed by the Uzbek startup MindBridge AI. UzbekVoice AI can understand and respond to voice commands in Uzbek and provides a wide range of services such as news updates, weather forecasts, and navigation assistance. The system utilizes deep learning models to process speech and generate responses in Uzbek. UzbekVoice AI has the potential to improve the accessibility of digital services in Uzbekistan by providing a convenient and user-friendly interface for Uzbek speakers.

MohirAI is another innovative application of NLP AI tools in Uzbekistan. MohirAI is an AI-powered chatbot developed by the Uzbek startup MohirTech. The chatbot is designed to provide personalized financial advice to users in Uzbek, based on their financial goals and risk tolerance. MohirAI utilizes advanced NLP AI techniques, including intent recognition and entity extraction, to understand users' queries and provide relevant responses. MohirAI has the potential to improve financial literacy and access to financial services in Uzbekistan, where there is a significant unbanked population. The success of UzbekVoice AI and MohirAI highlights the potential of NLP AI tools for developing innovative applications in Uzbekistan. However, there are still significant challenges to overcome. One of the significant challenges is the limited digital footprint of Uzbek language content. Developing large-scale annotated datasets and training deep learning models for Uzbek requires significant resources and expertise. Furthermore, there is a shortage of qualified experts in the field of NLP AI in Uzbekistan. To overcome these challenges, there is a need for increased research and development in Uzbek NLP AI tools. This can be achieved through collaboration between industry, academia, and government institutions. Furthermore, there is a need for the development of training programs and educational initiatives to train the next generation of NLP AI experts in Uzbekistan. Recent advancements in Uzbek NLP AI tools offer significant potential for developing innovative applications in Uzbekistan. UzbekVoice AI and MohirAI are two notable examples of the successful implementation of NLP AI tools in Uzbekistan. However, there are still significant challenges to overcome, such as limited linguistic resources and a shortage of qualified experts. To realize the full potential of NLP AI tools in Uzbekistan, there is a need for increased research and development, collaboration, and education initiatives.

Conclusion.

Recent advancements in NLP AI tools have opened up new opportunities for innovation and development in Uzbekistan. UzbekVoice AI and MohirAI are two examples of how NLP AI tools are being successfully implemented in Uzbekistan to improve accessibility and financial literacy. However, there are still challenges to overcome, such as the limited digital footprint of the Uzbek language and a shortage of qualified experts in the field of NLP AI. To address these challenges, increased research and development, collaboration, and education initiatives are needed to realize the full potential of NLP AI tools in Uzbekistan. By investing in the development of NLP AI tools, Uzbekistan can improve access to digital services and financial literacy, create new job opportunities, and drive economic growth.

References:

1. Alpana. "Challenges with Chatbots - not just technical," Medium. 2017.
2. Boltayev, B., Xujayev, O. " Building artificial intelligence chatbot, with help of distributed data analysis". Thematic Journal of Applied Sciences. 2022.
3. Mukhamadiyev, A., et al. "Automatic speech recognition method based on deep learning approaches for uzbek language". Sensors. 2022.
4. Sayfitdinov U . " Bot-Based Researches In Uzbekistan" European Journal of Research Development and Sustainability. *EJRDS*.2021.
5. MohirTech. (n.d.). MohirAI. Retrieved from <https://www.mohir.ai/>

BILINGVIZM TURLARI VA IKKINCHI TIL LEKSIK BAZASI**Israilova Saodat**

O'zbekiston Milliy universiteti katta o'qituvchisi, PhD

Ergasheva Ma'suda

O'zbekiston Milliy universiteti magistranti

Annotatsiya: Jahon tilshunosligida til hodisalarini boshqa ijtimoiy hodisalar bilan bog'liq holda yorituvchi lingvistik tadqiqotlar XX asr boshlarida Fransiya, Rossiya, Chexiya, AQSh kabi mamlakatlarda vujudga kela boshladi. Tilni ijtimoiy institutlar va jamiyat evolyutsiyasi bilan birga o'rganish, til va jamiyat, lingvistik vaziyat, ijtimoiy va kommunikativ tizim, kommunikativ kompetensiya, lingvistik kod, kodni o'zgartirish, ikkitillilik, ko'ptillilik, til siyosati, adabiy til, milliy tillar, til taraqqiyoti modellari masalalari olimlarning diqqat markazida bo'ldi.

Kalit so'zlar: bilingvizm, ikki tillilik, kognitiv tilshunoslik, tabiiy til, koordinativ, psixolingvistik.

Annotation: In world linguistics, linguistic studies that shed light on language phenomena in connection with other social phenomena began to appear in countries such as France, Russia, the Czech Republic, and the United States at the beginning of the 20th century. Study of language along with social institutions and society evolution, language and society, linguistic situation, social and communicative system, communicative competence, linguistic code, code switching, bilingualism, multilingualism, language policy, literary language, national languages, the issues of models of language development have become the focus of attention of scientists.

Keywords: bilingualism, bilingualism, cognitive linguistics, natural language, coordinative, psycholinguistics.

Ko'p tillilik masalasi bilan bog'liq g'oyalar XX asr boshlarida yetuk tilshunoslar tomonidan ilgari surilib, bu boradagi qarashlar nazariy bayon qilingan. XX asrning 50-yillaridan keyin qayd etilgan jihatlarni o'rganish lingvistikaning asosiy masalalari sifatida e'tirof etilib, yirik tadqiqotlarning vujudga kelishiga zamin bo'ldi. XX asrning ikkinchi yarmida yuqoridagi masalalar yuzasidan qilingan tadqiqotlarda eksperimental tekshirishga o'tish, aniq faktlarni matematik asoslar vositasida tavsiflash, izchillik, ma'lumot to'plashga jiddiy e'tibor berish, faktlarni miqdoriy va statistik tahlil qilish kabi xususiyatlar ustunlik qildi. XXI asrga kelib, til siyosati muammolarini ilmiy-nazariy va amaliy o'rganish dunyo tilshunosligida keng tus oldi. Globallashuv jarayonida dunyodagi millatlar, xalqlar, jamoalarning muloqotini kuzatish, taraqqiyotning tilga ta'siri, tilning rivojlanishga ta'siri masalalarini yoritish, sivilizatsiya jarayonida olamning lisoniy manzarasi (ko'rinishi), Yer shari aholisi o'rtasida ko'p ishlatilayotgan tillar va ularning keng qo'llanilish omillari, bilingvizm, pluralingvizmning lingvistik xususiyatlarini ilmiy tadqiq etish katta ahamiyat kasb etmoqda. Mustaqillik yillarida O'zbekistonning til siyosati, mamlakatdagi lingvistik vaziyat, til siyosati modellari tadqiqi, bilingvizm va multilingvizm hodisalari yuzasidan qator ishlar amalga oshirilgan.

Bu ijtimoiy talab yoki biror sabab bilan yuzaga keladi. Uning tabiiy vujudga kelishiga bir davlat hududida ikki va undan ortiq xalq vakillarining jips yashashi ham sabab bo'ladi.

Bilingvist odam miyasi bir nechta sezilarli afzalliklarga ega bo'ladi. Ba'zilarini hattoki ko'rish mumkin: kulrang moddaning ko'proq zichlikka ega bo'lishidir. Bundan tashqari, ikkinchi tilni ishlatganda miyaning ba'zi bir qismlarining faollanishining ko'payishi kuzatiladi. Buni miyani charxlash, deb aytishimiz mumkin.

Hozirgi kunda bilingvizmning kognitiv afzalligi juda tushunarli ko'rinadi. Ammo o'z vaqtida bu ekspertlarni lol qoldirgan. XX asrning 60-yillarigacha ikki tillilik, tillar orasida doimiy o'tishga kerak bo'lgan energiyani sarflagani uchun, bola rivojlanishni sekinlashtiradigan to'siq deb hisoblangan. Ushbu nuqtayi nazar, birinchi navbatda, noto'g'ri bo'lgan ilmiy ishlarga asoslangan edi. Oxirgi paytdagi tadqiqotlar bir necha til bilish insonning qaror qabul qilish, vazifalar o'rtasida o'tish, e'tiborni jamlash kabi fazilatlarini rivojlanishiga olib kelishini tasdiqlab berdi.

Tabiiy bilingvizm sun'iy bilingvizmdan farqli ravishda o'rganilmaydi, balki egallanadi. Tabiiy ikki tillilik, ko'pincha ikki tilli oilalarda, shuningdek, ikki tilli kichik ijtimoiy guruhlarda, masalan, umumiy bolalar guruhlarida sodir bo'ladi, bu yerda bolalar tengdoshlari bilan bevosita muloqotda ikkinchi tilni egallaydilar. Tilni egallash bir vaqtning o'zida uni o'rganish bilan qo'llab-quvvatlanishi mumkin. Tabiiy bilingvizmda bir tildan ikkinchi tilga o'tish avtomatik tarzda kuzatiladi. Bilingvizmning sun'iy turi – ikkinchi til notabiiy usulda (tilni shu til egalari yoki o'qituvchi bilan muloqot qilish orqali o'zlashtiradi) va ikki tildan doimiy ravishda foydalanilmaydi.

Inson balog'atga yetgandan keyin yoki balog'at yoshida ikkinchi tilni o'zlashtirganda ikkinchi til ona tili asosida o'rganiladi. Bilingvizmning bunday turi tilshunoslar tomonidan "subordinativ" dan farqli ravishda "koordinativ" deb ataladi. Bilingvizmning "koordinativ" turi, ko'pincha, parallel bilingvizm tufayli bilan yuzaga keladi. Uning o'ziga xos xususiyati shundaki, turli tillardagi so'zlar va ularning ma'nolari o'zlashtiriladi va xotirada bir-biridan alohida saqlanadi. "Subordinative" bilingvizmda ikkala tildagi ma'no birliklari mutlaqo bir xil, chunki ular ona tili orqali qabul qilingan bo'ladi [1, p. 120]. Ikkinchi tilning grammatik va morfologik tuzilmalarini o'zlashtirish ham ona tilining tegishli tuzilmalari orqali amalga oshiriladi. Shuning uchun interferensiya hodisasi, ko'pincha, ikkinchi tilni o'rganish jarayonida barcha til sathlari: fonologiyada, semantikada, grammatikada sodir bo'ladi.

Ushbu ta'riflar aqliy leksika muammolariga kamida ikkita yondashuvni ta'kidlashga imkon beradi: u inson xotirasidagi lug'at manbai (leksik bilimlar ombori sifatida) va nutqni idrok etish va ishlab chiqarish jarayonida axborotni qayta ishlash mexanizmlari murakkab ko'p bosqichli model sifatida tushuniladi. Nutqni ishlab chiqarish jarayonida inson doimiy ravishda o'zining aqliy lug'atidan foydalanadi, o'z fikrini shakllantiradi va muayyan so'zlar va leksik birliklarni tanlaydi. Psixik leksikaning tuzilishida konseptual, semantik-leksik va fonologik darajalar mavjud. Ayrim so'zlar, iboralar va jumalarni hosil qilish jarayonida ushbu darajalarning faollashishi qonuniyatlari va xususiyatlari dolzarb masala bo'lib qolmoqda.

Olimlar orasida aqliy leksika qanday ishlashi haqida yagona fikr yo'q. Garchi ko'pchilik olimlar uning turli xil aqliy tarmoqlar: semantik, sintagmatik va tematik guruhlar, sohalar va so'z turkumlari shaklida tashkil etilganiga ishonishsalar ham tadqiqotlar shuni ko'rsatdiki, agar so'z yoki leksik obyekt turli subtarmoqlarga kiritilgan bo'lsa, bu ushbu so'zni xotiradan keyingi faollashtirish va unga kirishni osonlashtiradi.

Yangi tilni rivojlantirish jarayonida shakllanadigan ikki tilli psixik leksikani o'rganish alohida o'rin tutadi. Ikki tilli leksikani modellashtirishda chet tilining o'zlashtirilgan leksik birliklari mental leksikada alohida kichik tarmoqlarni tashkil etishi eng asosli taxmindir. Biroq ona tillari va chet tillari doimo o'zaro ta'sirda bo'ladi va bu o'zaro ta'sir jarayonida ikki tilli psixik leksika shakllanadi. Yangi lug'atni o'rganishda avval o'rganilgan materiallarga murojaat qilinadi. Bu nuqtayi nazar chet tilini o'zlashtirish jarayonida chet el lug'ati qanday o'zlashtirilganligi, o'rganuvchi qanday qiyinchiliklarga duch kelishi va qanday leksik va fonetik xatolarga yo'l qo'yishini o'rganadigan tadqiqotlar bilan tasdiqlanadi. [9 p. 39]. Bolalar ikkinchi tilni kattalarga qaraganda osonroq o'rganishlari mumkinligi ilmiy adabiyotlarda keng tarqalgan. Voyaga yetganida ikkinchi (xorijiy) tilni o'rganish ko'pchilik tomonidan uzoq va mashaqqatli jarayon sifatida qabul qilinadi. Bir qator nazariy asoslar (birinchi navbatda, kognitiv-psixologik va affektiv), ikkinchi tilni o'rganish jarayonida bolalarning kattalardan ustunligi haqidagi tezisni qo'llab quvvatlaydi. Bolalar til o'rganish uchun alohida qobiliyatga ega.

Ikkinchi muhim jihat shundaki, bolalar dunyoni bilish jarayonida tilni o'rganadilar, ularga atrof-muhitni tushunish uchun til kerak. Bu holda ikkinchi til o'z-o'zidan o'zlashtiriladi, chunki bolalar uni o'rganishga e'tibor bermaydilar. Bolalardan farqli o'laroq, kattalar tilni maqsadli va ongli ravishda o'rganadilar, undan butunlay boshqacha maqsadda o'rganish uchun emas, balki tasvirlash uchun foydalanadilar [10, p. 41].

O'zbek mumtoz adabiyotining ko'pgina vakillari ham o'zbek, ham fors tilini yaxshi bilganlar. Masalan, Alisher Navoiy o'z ona tili — turkiy til bilan bir qatorda, fors tilini ham juda yaxshi bilgan. Bu tildagi asarlarini to'plab devon ("Devoni Foni") tuzgan. Bu devon forsiy til shoirlar, xususan, Jomiy tomonidan yuksak baholangan.

Ikki tilda ijod qiladigan adiblar hozir ham uchraydi. Jumladan, mashhur qirg'iz yozuvchisi Chingiz Aytmatov qirg'iz va rus tillarida erkin ijod qiladigan yozuvchidir. Umuman olganda, ikki tillilik ijtimoiy hodisa sifatida tillarning boyishi, rivojlanishi, umumbashariy madaniy saviyaning oshishiga muhim ta'sir etuvchi omildir. Tillarning bir-biriga har qanday ta'sir ko'rsatishi bilingvist odamlar bo'lishini talab etadi.

Xulosa qilib aytganda, bilingviallik shaxs ongida dunyoning adekvat tasviriga erishishga yordam beradi. Bilingvizm ijtimoiy va siyosiy hayotda muloqot qilish, jamoada ishlash, kommunikatsiya; xalqlar o'rtasida keng madaniy almashinuvni yoyishga hissa qo'shadi; turli millat vakillarining muloqotida til to'siqlarini minimallashtirish, xalqlarning o'zaro izolyatsiyasi ehtimolini istisno qiluvchi vosita sifatida namoyon bo'ladi; millatlararo muloqot tilining keng tarqalishiga hissa qo'shadi; umumiy va birlashgan jarayonda ishtirok etuvchi har bir tilning lug'at boyligini o'zaro boyitish va o'zaro ta'sir qilish, to'ldirish va kengaytirishga ham hissa qo'sha

Foydalanilgan adabiyotlar:

1. Zalevskaya A. A., Medvedeva I. A. Psixolingvisticheskie problemy uchebnogo dvuyazychiya. – Tver, 2002.
2. Verещagin Ye. M. Psixologicheskaya i metodicheskaya xarakteristika dvuyazychiya (bilingvizma). – M., 1969.
3. Döpke S. One parent –none language. An international approach / S. Döpke. – Amsterdam: Benjamin's, 1992.

4. Schwarz M. Kognitivismus und Lexikon / M. Schwarz // Die Ordnung der Wörter: kognitive und lexikalische Strukturen / Hrgs. Von Gisela Harras. – Berlin; New York: Walter de Gruyter, 1995.
5. Raupach M. Das mehrsprachige mentale Lexikon / M. Raupach // In: Börner/Vogel (Hrsg.). – 1997.
6. Plieger P. Struktur und Erwerb des bilingualen Lexikons. Konzepte für die medienge- stützte Wortschatzarbeit /

KOMPYUTER LINGVISTIKASINING NUTQ BIRLIKLARI ANALIZATORLARI VA ULARNING TAHLILI

Bahriddinova Yulduzxon Bahriddin qizi

Termiz davlat universiteti

Xorijiy filologiya fakulteti o'qituvchisi

E-mail: yulduz.bahriddinqizi@gmail.com

Annotatsiya. Hozirgi kundagi jadal rivojlanib borayotgan fan-texnika asrida jamiyatning har bir sohasi o'z taraqqiyot yo'lida har kuni bir yangilikni kashf etib bormoqda va bu jarayonning vujudga kelishi uzviy ravishda kompyuter texnologiyasi bilan bog'liq bo'lmoqda desak mubolag'a bo'lmaydi. Jahon tilshunosligida bundan qariyb yarim asr oldin kompyuter lingvistikasi yo'nalishining paydo bo'lishi, hamda undan foydalanishning avzalliklari ham shular jumlasidandir. Ushbu maqolada kompyuter lingvistikasining rivojlanish jarayoni, u haqidagi tilshunoslarning ta'riflari, tabiiy tilni qayta ishlash (NLP) jarayoni, nutq birliklarining grammatik, morfologik, sintaktik analizatorlari va ularning tahlili muhokama qilinadi.

Kalit so'zlar: kompyuter lingvistikasi, NLP, nutq birliklari, grammatik, morfologik, sintaktik analizatorlar.

Annotation. In the dynamically developing age of Science and technology of the present day, each sphere of society is discovering a novelty every day on the path of its development, and the emergence of this process is inextricably linked with computer technology, which cannot be exaggerated. In World linguistics, this includes the emergence of the direction of computer linguistics half a century ago, as well as the advantages of its use. This article discusses the process of development of computer linguistics, the definitions of linguists about it, the process of Natural Language Processing (NLP), grammatical, morphological, syntactic analyzers of speech units and their analysis.

Keywords: computer linguistics, NLP, speech units, grammatical, morphological, syntactic analyzers.

O'tgan asrning ikkinchi choragida dunyo sivilizatsiyasida tezkor globallashtirish jarayoni boshlandi va bu jarayonda xalqlar o'rtasida axborot almashish ham siyosiy, ham iqtisodiy va ko'plab boshqa sohalarada asosiy zarurat hisoblanib qoldi. Aynan shunday vaziyatda tilshunoslikning yangi kompyuter lingvistika yo'nalishi ma'lum bir axborotni bir tildan ikkinchi tilga tarjima qilish, yoki matnni tezkor tahrirlash va tahlil qilish jarayoni bilan bog'liq ravishda yangi soha sifatida paydo bo'ldi va insoniyat hayotida axborot integratsiyalashuvi vujudga kela boshladi. Kompyuter lingvistikasi nazariyasi va amaliyoti, kognitiv psixologiya, matematika, informatika va falsafa bilan uzviy bog'liqdir. Kompyuter lingvistikasi qayd etilgan sohalarga oid masalalarni ham qamrab oladi. Kompyuter lingvistikasini o'rganish tilning o'ziga xos tizimini, ramzlar, belgilar, modellar ko'rinishidagi qiymatini, ijtimoiy funksiyalarini anglatishga, yangi informatsion texnologiyalar bilan tanishtirishga, tilshunoslikning nazariy masalalari, kognitiv

psixologiya, mantiq, falsafa, matematika bilan bog'liq qonuniyatlarini o'zlashtirishga xizmat qiladi.³

Kompyuter lingvistikasi dastavval sun'iy intellekt sohasida guruhlangan, ammo sun'iy intellekt rivojlanishidan oldin mavjud bo'lgan. Kompyuter lingvistikasi 1950-yillarda Qo'shma Shtatlarda chet tillaridan, xususan, rus ilmiy jurnallaridan matnlarni avtomatik ravishda ingliz tiliga tarjima qilish uchun kompyuterlardan foydalanish harakatlaridan kelib chiqqan.⁴ Kompyuterlar arifmetik (tizimli) hisob-kitoblarni odamlarga qaraganda ancha tezroq va aniqroq amalga oshirishi mumkinligi sababli, ular tilni qayta ishlashni boshlashlari uchun bu qisqa vaqt masalasi deb hisoblangan. Hisoblash va kvantitativ usullar tarixiy jihatdan zamonaviy tillarning oldingi shakllarini qayta tiklashga urinishda va zamonaviy tillarni til oilalariga kichik guruhlashda ham qo'llaniladi. Kompyuter lingvistikasida tabiiy tilni qayta ishlash jarayoni (Natural language processing) mavjud bo'lib, u kompyuter dasturining inson tilini og'zaki va yozma ravishda tushunish qobiliyati tabiiy til deb ataladi. Bu sun'iy intellektning (Artificial Intelligence) tarkibiy qismidir. NLP kompyuterlarga odamlar kabi tabiiy tilni tushunishga imkon beradi. Til og'zaki yoki yozma bo'ladimi, tabiiy tilni qayta ishlash sun'iy intellektdan foydalanib, haqiqiy ma'lumotni olish, uni qayta ishlash va kompyuter tushunadigan tarzda tushunish uchun foydalanadi. Odamlarda turli xil sezish organlari eshitish uchun quloq va ko'rish uchun ko'z bo'lgani kabi, kompyuterlarda o'qish uchun dasturlar va audio yig'ish uchun mikrofonlar mavjud. Odamlarda bu jarayonni qayta ishlash uchun miya bo'lgani kabi, kompyuterlarda ham tegishli jarayonlarni qayta ishlash dasturi mavjud. Tabiiy tilni qayta ishlash jarayonida, albatta matndagi so'zlar har jihatdan lingvistik analiz ya'ni tahlilga olinadi.

Taniqli tilshunos olim, professor B. Mengliyev o'zining kompyuter lingvistikasiga oid ilmiy izlanishlarida NLP da til va nutq birliklarining bir nechta nutq analizatorlari haqida so'z yuritgan. Xususan, lingvistik analizator avtomatlashgan qidiruv tizimida tabiiy tilni qayta ishlash jarayonida so'zning morfologik, sintaktik va semantik xususiyatini ifodalaydigan vosita yoki dastur modullari to'plami bo'lib, matndagi bir qator leksema(so'z)ning ketma-ketligi, gapning mazmunini tahlil qiluvchi manba, shu bilan birga, morfologik, sintaktik, semantik (ba'zida pragmatik) xususiyatlarni aniqlaydigan ko'p darajali gap tahlillaridan foydalanuvchi yaxlit tahlil dasturi deb ta'riflaydi.⁵

Lingvistik analizatorlar matnni qayta ishlash jarayonida matndagi har bir nutq birliklarini morfologik, sintaktik va semantik jihatdan tahlil qilishda muhim rol o'ynaydi va shunga ko'ra quyidagi turlarga bo'linadi:

- Morfologik analizator
- Sintaktik analizator
- Semantik analizator

Morfologik analizator – bu kirish so'zining morfologiyasini tahlil qilish dasturi, u har qanday matnning morfemalarini aniqlaydi. Uning asosiy vazifasi berilgan matndagi har bir so'zning qaysi so'z turkumiga mansub ekanligini avtomatik tarzda aniqlashdan iborat.

³ Xolmanova Z. T. Kompyuter Lingvistikasi 5120100 - Filologiya va tillarni o'qitish (o'zbek tili) yo'nalishi uchun. "Asian Book House". - Toshkent, 2020. – B 8

⁴ Хорошилов А.А. Теоретические Основы и методы построения систем Фразеологического машинного перевода. специальность - 05.13.17, Теоретические основы информатики Автореферат дис. на соискание ученой степени док. тех. Наук Москва 2006 –С 4

⁵ Mengliyev B. O'zbek tili lingvistik analizatorlarini yaratish. Ma'naviyat, maktabdan tashqari ta'lim, T, 2022
<http://marifat.uz/ruknlar/manaviyat/o-zbek-tili-lingvistik-analizatorlarini-yaratish.htm>

Sintaktik analizator matnning tabiiy til matnining sintaktik tuzilishini avtomatik tahlil qiladi, ayniqsa sintaktik munosabatlar va tarkibiy qismlarni jadval asosida ko'rsatib beradi. Sintaktik analizator qilish kompyuter lingvistikasi va tabiiy tilni qayta ishlash (NLP)ning muhim vazifalardan biridir va XX asrning o'rtalaridan boshlab kompyuterlarning paydo bo'lishi bilan u ham ma'lum bir ma'noda tadqiqot mavzusi bo'lib kelgan.

Semantik analizatorning asosini semantik tahlil va tavsif tashkil etadi. Semantik tavsif leksik noaniqlik darajasini kamaytirish va uni butunlay olib tashlash imkonini beradi. Sintaktik jihatdan bog'liq so'zlar noaniq bo'lsa, algoritm ularga berilgan semantik xususiyatlarning qiymatlarini solishtirish uchun barcha tegishli kombinator variantlarni ketma-ket izlashni amalga oshiradi. Ushbu jarayon ma'lum g'oyaga asoslangan bo'lib, bunday so'zlar ziddiyatli semalarga ega bo'lmasligi kerak. Semantik tavsifning asosiy maqsadi – leksik va grammatik omonimiyani olib tashlash yoki sintaktik tahlil natijasida olingan variantlar sonini kamaytirishdan iborat.⁶

Xulosa o'rnida shuni aytishimiz mumkinki, kompyuter tilshunosligi va uning rivojlanishi zamonaviy tilshunoslikning dolzarb masalasi va bunda qililayotgan har bir izlanish va ilmiy ishlar unda vujudga keladigan muammolarga yechim vazifasini bajarmog'i kerak.

Foydalanilgan adabiyotlar:

1. Mengliyev B. O'zbek tili lingvistik analizatorlarini yaratish. Ma'naviyat, maktabdan tashqari ta'lim, T, 2022
2. Xolmanova Z. T. Kompyuter Lingvistikasi 5120100 - Filologiya va tillarni o'qitish (o'zbek tili) yo'nalishi uchun. "Asian Book House". - Toshkent, 2020.
3. Хорошилов А.А. Теоретические Основы и методы построения систем Фразеологического машинного перевода. специальность - 05.13.17, Теоретические основы информатики Автореферат дис. на соискание ученой степени док. тех. наук Москва 2006.

⁶ Qarang: O'zbek tili lingvistik analizatorlarini yaratish. Ma'naviyat, maktabdan tashqari ta'lim, T, 2022
<http://marifat.uz/marifat/ruknlar/manaviyat/o-zbek-tili-lingvistik-analizatorlarini-yaratish.htm>

O'ZBEK-INGLIZ TILLARI UCHUN PARALLEL KORPUSGA ASOSLANGAN MASHINA TARJIMASI TEXNOLOGIYASI

Shamsiyeva Gulshoda

O'zbekiston Milliy universiteti

II bosqich magistranti

Annotatsiya: Mashina tarjimasi deganda bir tabiiy tilni boshqa tabiiy tilga avtomatik ravishda tarjima qilish tushuniladi. Ushbu maqolada biz o'zbek tilidan ingliz tiliga statistik mashina tarjimasi algoritmini taklif qilamiz. Ingliz va o'zbek tillari tipologik jihatdan uzoq tillardir. Ingliz tili juda cheklangan morfologiyaga ega. O'zbek tili esa juda boy agglyutinatativ tildir. O'zbekcha so'z tuzilmalar, tarjima qilinganda ingliz tilidagi bir nechta so'zlarning to'liq iboralariga mos kelishi mumkin. Ushbu maqolada fraze-baza modelidan foydalangan holda o'zbekchadan inglizchaga statistik mashina tarjimasi algoritmiga erishishni taklif qilingan. Bundan tashqari, statistik mashina tarjimasiga erishish uchun biz inglizcha-o'zbekcha korpusni rivojlantirishimiz kerak. Ushbu maqolada inglizcha-o'zbekcha korpusni rivojlantirish haqida ham qisqacha ma'lumot berilgan.

Kalit so'zlar: Mashina tarjimasi, tabiiy til, statistik mashina tarjimasi, korpus

Annotation: The machine translation means is a translating one natural language to another natural language automatically. English and Uzbek are typologically distant languages. The English language has very limited morphology and Uzbek is an agglutinative language with a very rich and productive derivational and inflectional morphology. The Uzbek word structures that can correspond to complete phrases of several words in English when translated. In this paper, propose that will achieve Uzbek to English statistical machine translation algorithm using phrase-base model. Moreover, in order to achieve statistical machine translation we need to develop English-Uzbek corpora. In this paper, we present briefly about English-Uzbek corpora development.

Keywords: Machine translation, natural language, statistical machine translation, corpora.

Аннотация: Средства машинного перевода — это автоматический перевод одного естественного языка на другой естественный язык. Английский и узбекский - типологически далекие языки. Английский язык имеет очень ограниченную морфологию, а узбекский язык является агглютинативным языком с очень богатой и продуктивной словообразовательной и флективной морфологией. Структуры узбекских слов, которые при переводе могут соответствовать полным словосочетаниям из нескольких слов на английском языке. В этой статье предлагается реализовать алгоритм статистического машинного перевода с узбекского на английский с использованием модели на основе фраз. Кроме того, для достижения статистического машинного перевода нам необходимо разработать англо-узбекские корпуса. В этой статье мы кратко расскажем о развитии англо-узбекских корпусов.

Ключевые слова: машинный перевод, естественный язык, статистический машинный перевод, корпус.

Tipologik jihatdan ingliz va o'zbek tillari bir-biridan juda farq qiladi. Ingliz tilida morfologiya cheklangan va oddiy jumlar tartibi quyidagi tartibda: ega+kesim+to'ldiruvchi. O'zbek tili aglutinativ til bo'lib, juda boy va flektiv morfologiyaga ega bo'lib, gap tartibi odatda ega+to'ldiruvchi+kesimdir. Amaliy ahamiyatga ega bo'lgan yana bir masala shundaki, o'zbekchadan inglizchaga tarjima qilingan keng ko'lamli parallel matn resurslar yo'q. Statistik mashina tarjimasi algoritmini yaratishning dastlabki bosqichi parallel matnlarni yig'ish bo'lib, bu o'zbek va ingliz tillarining tarjimasi muhimdir. Bunday matnlarning manbalari esa ko'p emas.

Statistikaga asoslangan mashina tarjimasi texnologiyasining bir qancha usullari mavlud bolib, quyida ular haqida to'xtalib o'tamiz.

So'zga asoslangan tarjima usulida tarjimaning asosiy birligi tabiiy tillardagi so'zdir. [Abduraxmonova N., Tuliyeu.,2018;68]. Odatda, tarjima qilingan jumlar, qo'shma so'zlar asl jumladan farq qiladi. Masalan, ingliz tilidagi "happy" so'zini jumla kontekstiga qarab o'zbek tiliga "xursand" yoki "kayfiyati chog'" deb tarjima qilish mumkin. Oddiy so'zga asoslangan tarjima turli xil tuzilishga ega bo'lgan tillar o'rtasida tarjima qilishda qiyinchiliklarni keltirib chiqaradi. So'zga asoslangan tarjima tizimlari bitta so'zni bir nechta so'zlarga moslashtira olish asosida ishlaydi. Misol uchun, agar biz o'zbek tilidan ingliz tiliga tarjima qilgan bo'lsak, o'zbek tilidagi har bir so'z istalgan miqdordagi inglizcha so'zlarni hosil qilishi mumkin. Biroq, bitta o'zbekcha so'z hosil qiluvchi ikkita inglizcha so'zni guruhlashning iloji yo'q. IBM modellari, HMM modeli va Model 6 uchun o'quv dasturini o'z ichiga olgan GIZA++ paketi (GPLed) kabi so'zlarga asoslangan tarjima tizimlari mavjud. [Abduraxmonova N., Urdishev K., 2019; 131-7]. Bugungi kunda so'zga asoslangan tarjima modeli keng qo'llanilmaydi. Hozirgi vaqtda iboraga asoslangan tizimlar ko'proq qo'llaniladi.

Frazaga asoslangan tarjima usulining maqsadi so'zlar ketma-ketligini tarjima qilish orqali so'z asosidagi tarjima cheklovlarini kamaytirishdir, tarjima uzunligi har xil bo'lishi mumkin. [Abduraxmonova N., Urdishev K., 2019; 131-7]. Tarjima uchun tanlangan iboralar iboralar tarjimasi jadvali asosida birma-bir joylashtiriladi va keyin til tuzilishini yaxshilash uchun qayta tartiblanishi mumkin. Morfologik boy tillar uchun iboraga asoslangan model yaxshi natija beradi.

Til modeli statistik mashina tarjimasining zaruriy komponentidir. [Abduraxmonova N., Urdishev K., 2019; 131-7]. Til modeli tarjimini iloji boricha ravon qilishga yordam beradi. Til modeli – bu tarjima qilingan jumlaning oladigan va uning eng ravon versiyasi ehtimolini qaytaradigan funksiya. Yaxshi til modeli, masalan, "maktabdan kelayotgan yigitlar" jumlasiga "maktab o'quvchisi kelayotgan" gapiga nisbatan yuqoriroq ehtimollikni belgilaydi. Til modelining yana bir funktsiyasi shundaki, u so'z tanlashda ham yordam berishi mumkin. Agar xorijiy so'zda bir nechta ehtimolli tarjimalar bo'lsa, bu funktsiyalar maqsadli tildagi muayyan kontekstlarda yaxshiroq tarjima qilish ehtimolini beradi. [Kubedinova L., Khusainov A., Suleymanov D., Gilmullin R., Abdurakhmonova N., 2020; 90-101].

O'zbek tilidan ingliz tiliga statistik mashina tarjimasi algoritmiga erishish uchun biz iboraga asoslangan modelni qo'llaymiz. O'zbek-ingliz tillarini solishtirganda o'zbek tilidagi ba'zi so'zlar bir nechta inglizcha so'zlarga yoki aksincha tarjima qilinishi mumkin. Bunday hollarda so'zga asoslangan modellar samarasiz bo'ladi. O'zbek tilidagi kirish gap avval so'z birikmalariga bo'linadi, so'ngra har bir ibora inglizcha iboraga tarjima qilinadi. Nihoyat, iboralar qayta tartiblanadi. Shuni ta'kidlash kerakki, iboraga asoslangan statistik mashina tarjimasi modelida iboralarni tarjima qilish jadvali oldin kiritilgan ma'lumotlarga asoslanadi va oldindan belgilangan model qayta tartiblashni amalga oshiradi.

Statistik mashina tarjimasini ehtimollik modeli bo'lib, berilgan til yo'nalishi bo'yicha tarjima misollarining oldingi kuzatuvlari asosida natijalarni yaratishni o'rganadi. U parallel korpus sifatida ham tanilgan. Statistik mashina tarjimasini qilishning bir qancha afzalliklariga ega. Biroq, ishonchli tarjimani olish uchun statistik mashina tarjimasini yetarli miqdordagi parallel korpusni talab qiladi, u esa so'zlarni tushuntirish va ularning tarjimasini ta'minlaydi. Bu shuni anglatadiki, bizda so'zlarning ma'nosini tushuntiradigan lug'at bo'lishi kerak. Demak, statistik mashina tarjimasini yaratishdagi asosiy to'siq berilgan tillarning parallel korpuslarini ishlab chiqishdir.

Parallel korpusni rivojlantirish uchun bizga ma'lumotlar bazasi kerak bo'ladi. Ma'lumotlar turli kontekstdagi o'zbek-ingliz va inglizcha-o'zbekcha matnlardir. Ma'lumotni turli yo'llar bilan yig'ish mumkin. Biz professional tarjimonlar tomonidan electron .txt formatida tarjima qilingan ilmiy maqolalardan ma'lumotlarni yig'amiz. Ushbu maqolalar o'zbek tilidagi yuzdan ortiq ilmiy matnlarni ingliz tiliga professional tarjimonlar tomonidan tarjima qilingan holatidir. Ushbu ilmiy maqolalar tibbiyot, tarix, tarjima va siyosat kabi turli sohalarni qamrab oladi.

Maqolada biz o'zbek tilidan ingliz tiliga statistik mashina tarjimasini tizimi algoritmi haqida to'xtaldik. Statistik mashina tarjimasiga erishish uchun parallel korpus sifatida ma'lumotlarni yig'dik va shuning uchun ingliz-o'zbek korpusini rivojlantirishni taklif qildik. Bundan tashqari, statistik mashinalarni o'rganish usullarini muhokama qildik. Statistik mashina tarjimasiga erishishning bir necha usullari mavjud bo'lib, biz o'zbekchadan ingliz tiliga tarjima qilish uchun iboraga asoslangan usulni taklif qildik. Aggilyutativ tilni tarjima qilishda iboraga asoslangan usullar yaxshi natija berishini ko'rish mumkin.

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-Khorezmii 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. aforef." (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, 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.
6. 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;5(03):89-100.
7. 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
8. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. In Proceedings of the International Conference on Language Technologies for All (LT4All) 2019.
 9. A. Ismailov, M. M. A. Jalil, Z. Abdullah and N. H. A. Rahim, "A comparative study of stemming algorithms for use with the Uzbek language," 2016 " 3rd International Conference on Computer and Information Sciences (ICCOINS), 2016, pp. 7-12, doi: 10.1109/ICCOINS.2016.7783180.
 10. 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

TABIIY TILNI QAYTA ISHLASH JARAYONIDA YUZAGA KELADIGAN ASOSIY MUAMMOLAR

Alijonova Fayruza Tohirjon qizi

Mirzo Ulugʻbek nomidagi

Oʻzbekiston Milliy universiteti magistranti

Annotatsiya: Oʻzbek tili murakkab soʻz va grammatik shakllarga egaligi sababli Tabiiy tilni qayta ishlash (NLP) jarayonida dolzarb muammolar yuzaga keladi. Maqolada ayrim xatoliklar va ularni bartaraf etish ustida olib borilayotgan izlanishlar xususida soʻz yuritamiz.

Kalit soʻzlar: Tabiiy tilni qayta ishlash (NLP), sunʼiy intellekt, ibora, koʻchma maʼno, oʻz maʼno, omonimlar.

Abstract: Due to the fact that the Uzbek language has complex words and grammatical forms, actual problems arise in the Natural Language Processing (NLP) process. In the article, we will talk about some errors and the research that is being carried out to eliminate them.

Keywords: Natural language processing (NLP), artificial intelligence, phrase, figurative meaning, proper meaning, homonyms.

Аннотация: В связи с тем, что узбекский язык имеет сложные слова и грамматические формы, в процессе обработки естественного языка (NLP) возникают актуальные проблемы. В статье мы расскажем о некоторых ошибках и проводимых исследованиях по их устранению.

Ключевые слова: Обработка естественного языка (НЛП), искусственный интеллект, фраза, переносное значение, собственное значение, омонимы.

Sunʼiy intellekt bizning kundalik hayotimizning bir qismiga aylandi - Alexa va Siri, matn va elektron pochta avtomatik tuzatish, mijozlarga xizmat koʻrsatish chatbotlari. Ularning barchasi yozma va ogʻzaki nutqda inson tilini qayta ishlash, "tushunish" va javob berish uchun mashinani oʻrganish algoritmlari va Tabiiy tilni qayta ishlash (NLP) dan foydalanadi.

Tabiiy tilni qayta ishlash (NLP) quyidagi muhim vazifalarni bajaradi:

Nutqni aniqlash va analiz qilish;

Matnni tahlil qilish;

Axborot olish va axborotlarni izlash;

Gaplarni tahlil qilish;

Matn hissiyotini aniqlash;

Savol-javob tizimlarini toʻgʻri tashkil etish;

Matn yaratish.

Matnda ifodalanayotgan maʼno (ijobiy, salbiy, neytral) tabiiy tilni qayta ishlash jarayonida ayrim muammolar natijasida koʻplab chalkashliklarni keltirib chiqaradi. Tabiiy tilni tushunish (Natural language understanding) jarayonida jonli inson nutqini analiz qilish va ifodalanayotga konteksni aniqlashda murakkab muammolar yuzaga kelmoqda.

Quyida ulardan ayrimlariga toʻxtalib oʻtamiz.

NLP katta foyda keltiradigan kuchli vositadir, ammo tabiiy tilni qayta ishlashda hali ham bir qator cheklovlar va muammolar mavjud:

Ko'chma ma'noli so'zlar va iboralar;

Omonimlar;

Sinonimlar;

Ironiya va sarkazm;

Noaniqlik;

Matn yoki nutqdagi xatolar;

So'zlashuv tillari va jargon;

Domenga xos til;

Kam manbali tillar;

Tadqiqot va ishlanmalarning yetishmasligi;

Matn yoki nutqdagi noaniqliklar, dialektlar, jargon, sarkazm kabilar matnlarni tahlil qilish, nutqni tushunib unga mos javobni qaytarishda chalkashliklarni keltirib chiqaradi. Quyida biz ko'chma ma'noli so'zlar va omonimlar, tabiiy tilni qayta ishlash jarayonida ular bilan yuzaga keladigan muammolar va ularga qaratilgan yechimlar xususida fikr yuritamiz.

Ko'chma ma'noli so'zlar va iboralar

Tildagi so'zlar bir ma'noli va ko'p ma'noli bo'ladi. Gap ichida ham, gapdan tashqarida ham aynan bir ma'noni ifoda etuvchi o'zlar bir ma'noli so'zlar deyiladi. Masalan, kompas, marmar, morfologiya, taassurot, xulosa, gigiena, g'oya, mimika, qobiliyat, fonema va b.

So'zlar ba'zan alohida holatda, nutqdan tashqarida bir xil ma'noni, gap tarkibida boshqa so'zlar bilan bog'liq holda boshqa bir ma'no anglatishi mumkin. Masalan, pasaymoq so'zi gapdan tashqarida olinganda erga tomon yaqinlashmoq ma'nosini bildiradi. Gap tarkibida esa bilimi pasaydi, ko'rish sezgisi pasaydi kabi boshqa ma'nolarni bildiradi.

Bu so'zlar birdan ortiq leksik ma'no ifodalayotgani uchun ko'p ma'noli so'z sanaladi.

Google tarjimon dasturi yordamida o'zbek tilidan ingliz tiliga o'girilgan namunani keltirib o'tamiz.



Bu yerda tarjimon dastur ariqning labi birikmasini so'zma so'z lip of the stream shaklida tarjima qilyapti.

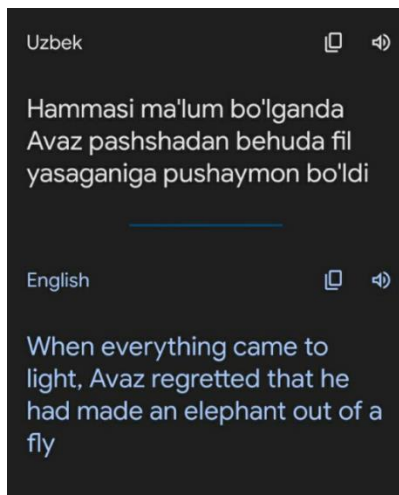
So'z yoki matnlarni tarjima qilishda iboralar ham jiddiy muammolarni yuzaga keltiradi.

Biror tushuncha yoki fikr ifodalov-chi, odatda, so'zlarning turg'un birikmasi (so'z birikmasi yoki gapga teng jumla) iboralar deyiladi.

- bir yoqadan bosh chiqarmoq – hamjihat bo'lmoq, yakdil bo'lmoq;

- burgaga achchiq qilib, koʻrpa kuydirmoq - arzimagan narsani deb, jahl ustida nojoʻya, zararli ish qilib qoʻymoq;
- pashshadan fil yasamoq – biror voqeani boʻrttirib koʻrsatmoq;
- qoʻy ogʻzidan choʻp olmagan – nihoyat darajada beozor, moʻmin

Quyidagi namunada iboralarni tarjima qilish jarayonida yuzaga kelgan kamchilikni koʻrishingiz mumkin



Namunada tarjimon dastur birikmani soʻzma soʻz tarjima qilganligini koʻrishimiz mumkin.

Maʼlumki, omonim soʻzlar shakli bir xil, ammo semantikasi turfa xil boʻlgan leksik birliklar hisoblanadi. Avtomatik qayta ishlashda omonimlik hodisasi quyidagi birliklarda mavjud: 1) oʻz omonimligi – shakldoshlik soʻz asosida boʻladi, yaʼni muayyan soʻz bir soʻz turkumi yoki bir necha turkumga mansub maʼno beradi. Masalan:

1. Avval b i r l i k l a r n i , keyin oʻnliklarni qoʻshamiz. (Ot)
2. Kuch – b i r l i k d a . (Ot)
3. Ogʻirlik b i r l i k l a r i . Til b i r l i k l a r i . (Ot)

1. Suv b o r joyda hayot bor. Ruchkang b o r m i ? (Modal)
2. Bir necha b o r taklif yubordim. (Hisob soʻz)
3. Bor – kimyoviy element. (Ot)
4. Ishga b o r m o q . Maktabga b o r m o q . (Fʼel)

Birinchi ustunda “birlik” soʻzi bir soʻz turkumi doirasida omonim hisoblanadi: 1) oʻngacha boʻlgan butun son; grammatik koʻplik aksi (ot); 2) birlashish, hamjihatlik (ot); 3) bir turdagi miqdorlarni oʻzaro baholash uchun qabul qilingan oʻlchov; til qurilishiga xos termin (ot). Ikkinchi ustunda “bor” soʻzi ikki xil turkumga mansub shakldoshlikni yuzaga keltirgan: 1) mavjud (modal); 2) marta, daʼfa, bora (hisob soʻz)

2. Qoʻshimcha omonimligi – muayyan qoʻshimcha vazifasiga koʻra qoʻshimchalarning turli guruhiga mansub boʻladi. Masalan:

-ki koʻchki, tepki, turtki (ot yasaydi);
ustki, ichki, kechki (sifat yasaydi);

(i)ng uying, kitobing, ishing (shakl yasaydi: sintaktik mun.shakli – egalik qoʻshimchasi);
koʻring, boring, tayyorlang (shakl yasaydi: sintaktik mun.shakli – shaxs-son qoʻshimchasi).

3. Ibora omonimligi – shakli, ya'ni tuzilishi bir xil, ammo semantikasi turlicha bo'lgan frazeologik birliklar.

uyni boshiga ko'tarmoq (qattiq shovqin soldi);

onasini boshiga ko'tarmoq (yuksak darajada izzat-hurmat qildi);

4. Gap omonimligi – muayyan gap ifoda maqsadi yoki mazmuniga ko'ra farqlanadi.

Ishni bajarmadim. (darak gap);

Ishni bajarmadim? (so'roq gap).

Rus tilida omonimlik turlarida so'zlarning turkumlari bo'yicha omonimlik, morfologik omonimlik va leksik omonimlik farqlanadi [2]. E'tiborli jihati shundaki so'zlarni morfologik va leksik omonimligi bo'yicha guruhlanishi tabiiy tilni qayta ishlashda muhim ahamiyat kasb etadi. Morfologik omonimlikda bir turkumga mansub bo'lgan shakldosh so'zlar lemma (asosi)si turlicha, ammo muayyan shakllaridagina omonimlikni yuzaga keltiruvchi so'zshakllar e'tiborga olinadi. Masalan:

lemmasi terim terim – mening terim (-im egalik qo'shimchasi: I shaxs, birlik);

lemmasi terim terim – hosil

Leksik omonimlikda bir lemmaga mansub so'z turli ma'nolarni beradi:

bog' – to'dalab bog'langan holat (ot);

Bog'lamoq, bog'lam. B i r b o g ' piyoz ot.

bog' – o'simlik va daraxtlar ko'p ekilgan joy (ot).

bog'-rog', bog'-bo'ston;

Uzumini ye, b o g ' i n i surishtirma.

Omonimshakllarning morfologik va leksik guruhlanishi matnlarni qayta ishlovchi dasturiy ta'minot va tizimlarda lemmatizatsiya va stemming jarayonlari uchun muhim sanaladi [3]. Har ikki texnologiya so'z yoki so'zshaklning asosini topishga yo'naltirilgan bo'lib, Ta'kidlash o'rinliki, omonimlik hodisasi avtomatik qayta ishlash jarayonida eng dolzarb masala hisoblanadi. Shu bois NLPda omonimshakllarni aniqlash va ularni tahlil qilish maxsus o'rganiladi, hatto bir necha usullar ham ishlab chiqilgan.⁷

Quyida tarjima jarayonida kuzatiladigan omonimlar bilan bog'liq xatoliklarni ko'rib chiqamiz.



⁷ Abjalova Manzura Abdurashetovna. Omonimiya va lingvistik tizimlarda omonimlarni aniqlash usullari. Oriental Renaissance: Innovative, educational, natural and social sciences

Smallpox soʻzi oʻzbek tilida chechak degan maʼnoni. Maʼlumki, chechak soʻzi kasallik va gul maʼnolarini bildiradi. Berilgan misolda chechak soʻzi gul maʼnosida kelgan. Tarjimada esa kasallik maʼnosidagi smallpox soʻzi qoʻllanmoqda. Bu kabi xatoliklar tarjima jarayonida koʻplab kuzatiladi.

Yuqorida keltirilgan koʻchma soʻzlar, iboralar va omonimlar bilan bogʻliq xatoliklarni bartaraf etish yoʻllari ustida tadqiqotlar olib borilmoqda. Nilufar Abduraxmanova va shogirdlari tomonidan toʻldirib borilayotgan <http://uzbekcorpus.uz/> dan foydalangan holda tabiiy tilni qayta ishlash jarayonida yuzaga kelayotgan muammolarga bosqichma bosqich yechimlar ishlab chiqilmoqda. Shu jumladan ushbu corpusda hozirda iboralar va omonimlar lugʻati ishga tushirildi. Keyingi vazifa ushbu lugʻatlardan foydalangan holda matnlar va tarjima jarayonida xatoliklarsiz ishlaydigan dasturni ishga tushirishdan iborat. Hozirda dastur uchun omonimlarning modellari yigʻilmoqda. Ushbu modellar yordamida dastur foydalanuvchi uchun zarur boʻlgan omonimlarni toʻgʻri analiz qilish va kerakli oʻrinda toʻgʻri qoʻllashga yordam beradi. Korpusda xatolarsiz ishlaydigan dasturning yaratilishi kelajakda mashina tarjimasi, chatboxlar, lugʻatlar va sunʼiy intellect rivoji uchun qoʻyilgan muhim qadamlardan boʻlishi shubhasiz.

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 VI International Conference “Modern Problems of Applied Mathematics and Information Technology - Al-Khorezmiy 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) il dis. aforef." (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, 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.
6. 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;5(03):89-100.
7. 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

8. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.
9. Abjalova Manzura Abdurashetovna. Omonimiya va lingvistik tizimlarda omonimlarni aniqlash usullari. Oriental Renaissance: Innovative, educational, natural and social sciences, November 2021

SENTIMENT TAHLILGA OID XORIJ TADQIQOTLARI

Raximov Xasanboy Komiljonovich

Andijon davlat universiteti

1-kurs tayanch doktoranti

Annotatsiya: Sentiment tahlil qilish tabiiy tilni qayta ishlash (NLP) doirasidagi asosiy tadqiqot sohasi bo'lib, u ijtimoiy tarmoqlar, yangiliklar maqolalari va mahsulot sharhlari kabi matn manbalaridan sub'ektiv ma'lumotlarni avtomatik ravishda aniqlash, ajratib olish va ifodalash uchun kompyuter lingvistikasidan foydalanishni o'z ichiga oladi. Sentiment tahlil qilishning maqsadi matnda ifodalangan hissiy munosabat yoki fikrni - uning ijobiy, salbiy yoki neytral ekanligini aniqlashdir.

Kalit so'zlar: Sentiment tahlil, xorijdagi tadqiqotlar, tabiiy tilni qayta ishlash (NLP), mashinada o'rganish algoritmlari, ijtimoiy tarmoqlar monitoring, madaniy va lingvistik farqlar, ko'p tilli hissiyotlarni tahlil qilish.

Annotation: Sentiment analysis is a major research area within Natural Language Processing (NLP) that studies the use of computational linguistics to automatically detect, extract, and represent subjective information from text sources such as social media, news articles, and product reviews. The purpose of sentiment analysis is to determine the emotional attitude or opinion expressed in the text - whether it is positive, negative or neutral.

Keywords: Sentiment analysis, research abroad, natural language processing (NLP), machine learning algorithms, social media monitoring, cultural and linguistic differences, multilingual sentiment analysis.

Аннотация: Сентимент анализ является основной областью исследований в области обработки естественного языка (NLP), в которой изучается использование компьютерной лингвистики для автоматической идентификации, извлечения и представления субъективной информации из текстовых источников, таких как социальные сети, новостные статьи и обзоры продуктов. Цель сентиментального анализа состоит в том, чтобы определить эмоциональное отношение или мнение, выраженное в тексте, будь то положительное, отрицательное или нейтральное.

Ключевые слова: анализ сентиментального, исследования за рубежом, обработка естественного языка (NLP), алгоритмы машинного обучения, мониторинг социальных сетей, культурные и языковые различия, многоязычный сентимента анализ.

So'nggi paytlarda xorijda sentiment tahlil qilish bo'yicha tadqiqotlarga qiziqish ortib bormoqda. Turli kontekst va tillarda sentiment tahlil qilishni o'rganish uchun bir nechta tadqiqotlar o'tkazildi. Bir misol, turizm sohasida olib borilgan tadqiqotlar, bu yerda onlayn sharhlar va fikr-mulohazalarni tahlil qilish uchun sentiment tahlil qo'llaniladi. Ushbu tadqiqot natijalari turistik xizmatlar sifatini oshirish strategiyalarini ishlab chiqishda ma'lumot berishi mumkin. Sentiment tahlil qilish matnli ma'lumotlardan sub'ektiv ma'lumotlarni olish jarayonidir. Sentiment tahlil qilishning asosiy maqsadi - ijobiy, salbiy yoki neytral kabi matnning asosiy hissiyotini yoki kayfiyatini aniqlashdir. So'nggi yillarda sentiment tahlil qilish biznes, marketing, siyosat va

ijtimoiy media kabi ko'plab sohalarda tobora ommalashib borayotgan tadqiqot mavzusiga aylandi. Tadqiqot metodologiyasi nuqtai nazaridan, sentiment tahlil qilish tadqiqoti odatda ikkita asosiy yondashuvga amal qildi: qoidalarga asoslangan usullar va mashinani o'rganishga asoslangan usullar. Qoidalarga asoslangan usullar ijobiy yoki salbiy hissiyot qiymatlari bilan izohlangan so'zlar va iboralarning oldindan belgilangan leksikasiga tayanadi, mashinani o'rganishga asoslangan usullar esa matnlarni avtomatik ravishda ijobiy, salbiy yoki neytralga tasniflash uchun izohli ma'lumotlar to'plamidan o'rganadigan algoritmlardan foydalanadi. Tuyg'ularni tahlil qilish bo'yicha tadqiqotlardagi asosiy qiyinchiliklaridan biri matnlardagi his-tuyg'ularni aniqlash va qo'llanilgan tilning muammolarini to'g'ri aniqlash qiyinligidir. Ushbu muammoni hal qilish uchun tadqiqotchilar neyron tarmoqlar, chuqur o'rganish usullari va xususiyatlarni tanlash usullari kabi turli usullarni o'rganishdi. Ilovalar nuqtai nazaridan, hissiyotlarni tahlil qilish ijtimoiy media monitoringi, brend obro'sini boshqarish, mahsulot sharhlari va reytinglarini tahlil qilish, siyosiy his-tuyg'ularni tahlil qilish va hatto sog'liqni saqlash monitoringi kabi turli sohalarda qo'llanilgan.

Sentiment tahlil qilish - bu matndan sub'ektiv ma'lumotlarni olish bilan shug'ullanadigan tabiiy tilni qayta ishlash (NLP) sohasi. Bu ma'lumotlar fikrlar, munosabat va his-tuyg'ularni o'z ichiga oladi. Sentiment tahlil qilish butun dunyo bo'ylab tadqiqotchilar, biznes va hukumatlar orasida qiziqish uyg'otmoqda. Ushbu maqolada biz xorijda, xususan, Qo'shma Shtatlar, Evropa va Osiyoda o'tkazilgan sentiment tahlil tadqiqotlarini o'rganishni maqsad qilganmiz.

Qo'shma Shtatlarda hissiyotlarni tahlil qilish marketing va siyosat sohalarida keng qo'llanilgan. Tadqiqotchilar ijtimoiy tarmoqlar, yangiliklar maqolalari va onlayn munozaralardagi his-tuyg'ularni tahlil qilish uchun turli xil mashina o'rganishga asoslangan algoritmlarni ishlab chiqdilar. Ushbu algoritmlar matnlarni ijobiy, salbiy va neytral his-tuyg'ularga ajratish uchun chuqur o'rganish, tabiiy tilni qayta ishlash va ma'lumotlarni qidirish kabi usullardan foydalanadi. Ular brend obro'sini kuzatish, siyosiy masalalar bo'yicha jamoatchilik fikrini kuzatish va hatto aksiyalar narxini bashorat qilish uchun ishlatilgan.

Evropada sentiment tahlil qilish bo'yicha tadqiqotlar tilga xos muammolarga qaratilgan. Tadqiqotchilar fransuz, nemis va golland kabi turli Yevropa tillari uchun hissiyot leksikonlari va modellarini ishlab chiqdilar. Ular, shuningdek, matnlar bir nechta tillarda yozilgan ko'p tilli sharoitlarda sentiment tahlil qilishdan foydalanishni o'rganishdi. Bundan tashqari, tadqiqotchilar ijtimoiy fanlarda, xususan, onlayn muloqot, siyosat va psixologiyani o'rganishda hissiyotlarni tahlil qilishdan foydalanishni o'rganishdi. Osiyoda sentiment tahlil qilish bo'yicha tadqiqotlar turli ijtimoiy va madaniy kontekstlarda qo'llanilgan. Tadqiqotchilar xitoy va yapon matnlarida his-tuyg'u va hissiyotlarni tahlil qilish usullarini ishlab chiqdilar. Ular, shuningdek, ijtimoiy sog'liqni saqlashda, xususan, ijtimoiy media postlaridan ruhiy salomatlik muammolarini aniqlashda sentiment analizdan foydalanishni o'rganishdi. Bundan tashqari, sentiment tahlil qilish tabiiy ofatlar paytida ijtimoiy media ma'lumotlarini tahlil qilish va ijtimoiy masalalar bo'yicha jamoatchilik fikrini kuzatish uchun ishlatilgan. Umuman olganda, xorijda his-tuyg'ularni tahlil qilish bo'yicha tadqiqotlar sezilarli yutuqlarga erishildi va turli sohalarda qo'llash soni ortib bormoqda. Biroq hali ham hal qilinishi kerak bo'lgan muammolar mavjud. Bularga sentiment tahlil qilish metodologiyasida standartlashtirishning yo'qligi va ko'p tilli sharoitlarda matnlarni tahlil qilish qiyinligi kiradi. Ushbu muammolarni hal qilish va his-tuyg'ularni tahlil qilishning aniqligi va qo'llanilishini yaxshilash uchun qo'shimcha tadqiqotlar talab etiladi.

Boshqa bir tarafdin nazar tashlasak sentiment tahlil bu shaxslarning ma'lum bir brend, mahsulot yoki xizmatga nisbatan sub'ektiv fikrlari, munosabati va hissiyotlari haqida tushunchaga

ega bo'lish jarayoni. Ushbu uslub foydalanuvchining sharhlari, yoki ijtimoiy tarmoqdagi xabarlarini matnini tahlil qilish, umumiy ohang, kontekst va his-tuyg'ularni aniqlashni o'z ichiga oladi. U korxonalar, hukumatlar va notijorat tashkilotlari uchun bozor tadqiqotlari, mijozlar ehtiyojini qondirish va qarorlar qabul qilish jarayonlarida muhim vositaga aylandi. Globallashuv va madaniyatlararo muloqotning ahamiyati ortib borishi bilan sentiment tahlil turli mamlakatlar va mintaqalardagi iste'molchilar va manfaatdor tomonlarning his-tuyg'ulari va munosabatlarini tushunishning qimmatli usuliga aylandi. Chet elda his-tuyg'ularni tahlil qilish bo'yicha tadqiqotlar turli mamlakatlardagi odamlarning ma'lum bir mavzu yoki brendga nisbatan kayfiyati va munosabati haqida ma'lumot to'plashni o'z ichiga oladi. Bunga so'rovlar, ijtimoiy media monitoringi va onlayn sharhlar kabi turli usullar orqali erishish mumkin. Keyin to'plangan ma'lumotlar aloqa strategiyalari, marketing kompaniyalari va biznes qarorlarini optimallashtirish uchun ishlatilishi mumkin bo'lgan tendentsiyalar, naqshlar va tushunchalarni aniqlash uchun tahlil qilinishi mumkin.

Sentiment tahlil qilish tabiiy tilni qayta ishlash (NLP) doirasidagi asosiy tadqiqot sohasi bo'lib, u ijtimoiy tarmoqlar, yangiliklar maqolalari va mahsulot sharhlari kabi matn manbalaridan subyektiv ma'lumotlarni avtomatik ravishda aniqlash, ajratib olish va ifodalash uchun hisoblash texnikasidan foydalanishni o'z ichiga oladi. sentiment tahlilning maqsadi matnda ifodalangan hissiy munosabat yoki fikrni - uning ijobiy, salbiy yoki neytral ekanligini aniqlashdir. So'nggi yillarda ijtimoiy media platformalarining ko'tarilishi va mahsulot va xizmatlarning muvaffaqiyatini aniqlashda mijozlarning fikr-mulohazalaridan keng foydalanish bilan sentiment tahlil qilish tobora ommalashib bormoqda. Korxonalar mijozlar fikr-mulohazalarini tahlil qilish va o'z mahsulotlari, xizmatlari va brend imidjini yaxshilash uchun sentiment tahlildan foydalanadilar. Xuddi shunday, tadqiqotchilar ijtimoiy va madaniy hodisalarni o'rganish uchun sotsiologiya, psixologiya va siyosatshunoslik kabi sohalarda hissiyotlarni tahlil qilishdan foydalanganlar. Sentiment analiz qilishning eng keng tarqalgan va to'g'ridan-to'g'ri yondashuvi leksikonga asoslangan tahlildir. Ushbu usulda hissiyot ballari alohida so'zlar yoki iboralarga ularning semantik yo'nalishiga qarab belgilanadi, odatda, ijobiy, salbiy yoki neytral deb tasniflanadi. Keyin hissiyot ballari yig'iladi, hujjatdagi barcha so'zlar uchun qo'yilgan ballar yig'iladi va umumiy ball matnning hissiyotini aniqlash uchun ishlatiladi.

Biroq nazorat ostida va nazoratsiz o'rganish, neyron tarmoqlar, yashirin Markov modellari va chuqur o'rganish kabi mashinani o'rganish usullari hissiyotlarni tahlil qilishning aniqligi va ishonchliligini oshirdi. Nazorat ostidagi o'rganishda naqshlarni aniqlash va tasniflash modelini yaratish uchun ma'lum bir etiketli ma'lumotlar to'plamida mashinani o'rganish modeli o'qitiladi. Nazoratsiz o'qitishda ma'lumotlar yorliqlanmaydi va ma'lumotlarning statistik xususiyatlariga asoslangan naqshlar topiladi. Konvolyutsion neyron tarmoqlari (CNN) va takroriy neyron tarmoqlari (RNN) kabi chuqur o'rganish modellari odatda sentiment tahlil qo'llaniladi va hissiyotlarni aniqlashda eng zamonaviy aniqlikka erishdi. His-tuyg'ularni tahlil qilish tadqiqotlari turli sohalarda sezilarli ta'sir ko'rsatdi. Masalan, reklama sanoati reklama kompaniyalarining jamoatchilikka ta'sirini o'lchash uchun his-tuyg'ularni tahlil qilish texnologiyasini kiritishda sezilarli yutuqlarga erishdi. Tibbiyot sanoatida shifokorlar bemorlarning fikr-mulohazalarini tahlil qilish va xizmatlarini yaxshilash uchun har qanday salbiy fikrlarga amal qilish uchun sentiment tahlil qilishdan foydalanishlari mumkin. Sentiment tahlil qilish, ayniqsa, modellarning to'g'riligiga nisbatan cheklovlersiz emas. Modellar matnlardagi his-tuyg'ularni aniqlash va tahlil qilishda uzoq yo'lni bosib o'tgan bo'lsa-da, ular hali ham sarkazm, istehzo va satirani tan olmasligi

mumkin, bu esa natijalarning ishonchliligiga ta'sir qiladi. Bundan tashqari, his-tuyg'ularni idrok etish turli madaniyatlarda farq qiladi, bu esa madaniyatlararo hissiyotlarni tahlil qilish modellarini ishlab chiqishni qiyinlashtiradi. Xulosa qilib aytganda, hissiyotlarni tahlil qilish turli sohalarga, jumladan, reklama, siyosat va tibbiyotga sezilarli ta'sir ko'rsatdi. Leksikaga asoslangan tahlil, mashinani o'rganish va chuqur o'rganish kabi NLP usullaridan foydalanish matnlardagi his-tuyg'ularni aniqlashni sezilarli darajada yaxshilagan. Biroq sentiment tahlil qilish modellari hali ham ba'zi cheklovlarga ega, bu ularning aniqligi va samaradorligini oshirish uchun qo'shimcha tadqiqotlarni talab qiladi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N (2021) Kompyuter lingvistikasi (darslik), Toshkent: Nodirabegim, (b - 400).
2. Abduraxmonova N., Raximov X.(2023) O'zbek tili sentiment analizining nazariy masalalari. Toshkent INTERNATIONAL JOURNAL OF WORD ART Word Art, VOLUME 6, ISSUE 1 (b 35-40)
3. Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.
4. Kim, S. M., & Hovy, E. (2004). Determining the sentiment of opinions. In COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics (pp. 1367-1373).
5. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.
6. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.

TIL VA USLUB MATN TAHRIRINING BIRINCHI BOSQICHI SIFATIDA

Taxirov Zair Taxirovich

O'zbekiston Milliy universiteti,

O'zbek filologiyasi fakulteti

Kompyuter lingvistikasi va amaliy
tilshunoslik kafedrası professori, f.f.n.

Annotatsiya: Mazkur maqola adabiy tahrir jarayonining birinchi bosqichida amalga oshiriladigan mikrotahrirda matnni tuzatishda asosiy e'tibor qaratiladigan til va uslub xususida. Unda tasviriy fanlar hamda ularning o'rganish obyektlari, shuningdek, uslubiyat, uning tildan farqi, muayyan til birliklari va kategoriyalaridan foydalanish haqida so'z boradi.

Kalit so'zlar: Tasviriy fanlar, birikma, kategoriya, uslubiyat, mantiqiy tahlil, nutqiy muloqot, nutq egasi, ommaviy muloqot, badiiy nutq, adabiy nutq.

Annotation: This article focuses on language and style, which is the main direction of text correction during micro-editing, which is carried out at the first stage of the literary editing process. It talks about the visual sciences and their objects of study, as well as the methodology, how it differs from language and the use of specific language units and categories.

Keywords: Visual sciences, combination, category, methodology, logical analysis, verbal communication, orator, public communication, artistic speech, literary speech.

Аннотация: В данной статье речь идет о языке и стиле, что является основным направлением исправления текста при микроредактировании, которое осуществляется на первом этапе процесса литературного редактирования. В нем рассказывается о визуальных науках и их объектах обучения, а также о методологии, ее отличии от языка и использовании конкретных языковых единиц и категорий.

Ключевые слова: Визуальные науки, сочетание, категория, методология, логический анализ, речевое общение, оратор, публичное общение, художественная речь, литературная речь.

Tasviriy fanlar umumiy bir nom bilan – til deb ataladi. Ilmiy adabiyotlarda tasviriy fanlar quyidagicha guruhlanadi: fonetika, leksikologiya, frazeologiya, so'z yasalishi, grammatika. Grammatika o'z navbatida ikki qismga ajraladi: morfologiya, sintaksis.

Tasviriy fanlarning o'rganish obyekti bo'lib quyidagilar hisoblanadi:

- 1) til tizimidagi birliklar va kategoriyalar: muayyan tilning fonetik, grammatik, leksik qurilishi (so'z yasalishi leksika va grammatika oralig'idagi o'rinda turadi);
- 2) til birliklari va kategoriyalarining shakliy hamda ma'naviy (ular nimani anglatishi) tomonlari;
- 3) til birliklarini sinflarga guruhlash, ancha yirik til kategoriyalari;
- 4) tildagi birikmalar va kategoriyalarning til tizimidagi o'zaro munosabati va aloqadorligi.

Yuqorida qayd etilgan (yana boshqa) fanlar muayyan tildagi birikmalar va kategoriyalar tarkibini o'rganadi va tahlil etadi. Boshqacha aytganda, bu fanlar tilda borini o'rganadi. Muayyan

til fikrini, his-hayajonni ifodalashda, ma'lumot yetkazishda qanday vositalarga ega ekanligini tadqiq etadi.

Tilshunoslar til egasi, tildan foydalanuvchilarning o'z fikrini anglatishi, tushunchani ifodalashi, umuman, ma'lumot, xabar yetkazish uchun qaratilgan til imkoniyatlarini o'rganadilar. Keng ma'noda olganda, ma'lumot (informatsiya)lar tavsifiga ko'ra ilmiy, ish yuritishga oid, tahliliy, his-hayajonli, obyektiv, taxminiy, ishonarli, to'qima bo'lishi mumkin.

Biz til deganda, asosan "hozirgi o'zbek tili"ni tasavvur etamiz. Avvalo shuni aytish kerakki, hozirgi o'zbek tili – shunday tilki, u adabiy nutq me'yorlari asosida yaratilgan matnlarda ifoda topgan. Til tarixida u ikki bosqichdan o'tib, hozir uchinchi bosqichdagi ko'rinishga ega. Hozirgi o'zbek adabiy nutqi XX asr birinchi choragidan shakllana boshlagan va avvalgi bosqichdagiga qaraganda millatning ijtimoiy-siyosiy shakllanishi jarayonida ham fonetik, ham leksik, ham grammatik jihatdan umummilliy adabiy nutqqa aylanadi.

XX asr 50–70-yillari hozirgi o'zbek adabiy nutqining yanada rivojlanish davri bo'ldi, desak mubolag'a bo'lmaydi. Bu davrda o'zbek tilshunosligi rivojiga Ayub G'ulomov, Mazluma Asqarova, Shavkat Rahmatullayev, Shonazar Shoabdurahmonov va b. katta hissa qo'shgan. Bular tasviriy tilshunoslik yo'nalishida ilmiy faoliyat ko'rsatgan. O'zbek tilshunosligi bu davrda garchi rus tilshunosligi ta'sirida rivojlanishni boshdan kechirgan bo'lsa-da, uning fonetikasi, leksikologiyasi, morfologiyasi, sintaksisida milliy tilga xos yangi-yangi qirralar ochildi.

Yuqorida nomlari keltirilgan tilshunoslar qatoriga keyingi to'lqin tilshunoslari (aslida Ayub G'ulom maktabining vakillari)dan bo'lmish Azim Hojiyev, Miraziz Mirtojiyevlar kirib kelganlar. Ular an'anaviy – tasviriy tilshunoslik yo'nalishida faoliyat boshlagan bo'lsalar-da, keyinchalik, tahliliy yo'nalishga o'tadilar va o'zbek tilshunosligida semik tahlilga asos soladilar. Buning natijasida tilshunoslik uslubi rivojlanadi. O'zbek adabiy nutqi uslubiylatining rivojlanishi o'z navbatida mantiqiy tahlilni yuzaga chiqaradi. Mantiqiy tahlil esa adabiy tahrir fanining yuzaga kelishi va takomillashuviga kuchli turtki bo'ladi.

Xo'sh, uslubiylat o'zi nima? Uni qanday tushinish kerak? Uni tildan farqi nimada?

Uslubiylat – nutqiy muloqot jarayonida (til tizimini tashkil etuvchi til birikmalari va kategoriyalari)dan foydalanish to'g'risidagi fan. Nutqiy muloqot og'zaki va yozma bo'ladi. Gap shundaki, uslubiylat haqida so'z borganda tildagi vositalardan foydalanish yoki til vositalarining muayyan vazifani bajarishligi tushuniladi. Har ikki ibora ayni bir ma'no ifodasiga ega.

Shuni alohida ta'kidlash kerakki, uslubni tadqiq etish muayyan til material, muayyan tilda tuzilgan matn asosida amalga oshiriladi.

Tildan foydalanishni quyidagicha izohlash mumkin:

- a) uslubiylat maqsadlarda;
- b) muayyan maqsad va vazifalarga ko'ra (siyosiy, ilmiy, ish yuritish, shaxsiy, ijtimoiy, ezgu, jinoiy, raddiya, buyruq, qahr-g'azab, ayblovlar va b. mavzuda) matnlar tuzishda;
- v) muayyan muomala sharoitida.

Bunda ommaviy muloqot sharoitlari ko'zda tutilmoqda (gazeta, radio, televideniya va b.), muloqot guruhlar (parlament, ma'ruzalar, sudlov ishi, yig'ilishlar, konferensiyalar, seminarlar va b.), shaxslararo muloqotlar, dialoglar, suhbatlar (do'stona so'zlashuvlar, kursdoshlarning gaplashishlari, oila davrasida, ulfatlar to'plangandagi gurunglashuv va b.). Gazeta o'qiydiganlar, radiotinglovchilar, televideniya ko'rsatuvlarini kuzatib boruvchilarning ijtimoiy, siyosiy, yosh jihatdan tarkibi, ularning madaniyatlilik darajasi, ma'lumotlilik, saviyasi, dunyoqarashi, qay

darajada hissiyotga berilganligi va boshqalar, shuningdek, jamiyatning, xalqning bugungi kundagi turmush darajasi va hokazolar hisobga olinadi.

Biz yuqorida uslubiyat tildagi vositalardan foydalanish, dedik. Qanday? Bu savolga tildagi vositalarning amalda qo'llanilishi deganda ayonlashadigan mazmun javob bo'ladi. Aslida "til vositalaridan amalda foydalanish" masalasiga "muayyan tildan foydalanuvchi - nutq egasi" nuqtayi nazaridan yondashish to'g'ri bo'ladi. Tildan amalda foydalanish vaziyati va jarayonida biz til birikmalarining qo'llanilishiga shakliy tomondan emas, balki mazmuniy (mantiqiy) tomondan nazar tashlaymiz, unda yuz berayotgan ma'naviy (semantik) o'zgarishni aniqlaymiz va baholaymiz, ekspressivlikni kuzatamiz. Natijada, ayrim til birliklari va / yoki guruhining vazifasi o'zgarishligi, grammatik shakl, konstruksiya (gap)lar, fonemalar o'zgarishini kuzatamiz.

Agar tasviriy fanlarda til birikmalari, kategoriyalari adabiy nutq me'yorlariga mosmi, degan savolga javob berishi kerak bo'lsa, uslubiyatda esa qanday degan savolga, ya'ni muayyan til birliklari va kategoriyalaridan "qanday foydalaniladi..." ga javob beradi. Bunday savol qo'yilishi tabiiy, chunki uslubiyat muayyan til birliklari, kategoriyalari fikrni, his-hayajonni ifodalashda, fikr almashishda, xabar, ma'lumot yetkazishda qo'llanilishini o'rganadi va o'rgatadi. U tilning harakatlanishini kuzatadi.

Shunday qilib, uslubiyat tasviriy fanlardan farqli o'laroq tildan, til birikmalaridan nutqiy muloqot (kommunikatsiya)da foydalanishini, ularning amal qilishini tadqiq etadi. Shunga ko'ra, u tilshunoslikning funksional-kommunikativ – nutq madaniyati, oftologiya (to'g'ri nutq), til me'yorlari haqidagi, sotsiolingvistika, psixolingvistika, adabiy nutq me'yorlari kabi fanlari qatorida turadi.

Rus tilshunos olimi V.V. Vinogradov uslubiyatga ta'rif berib shunday deydi: "Tildan, ayniqsa, adabiy nutqdan foydalanishning barcha usullari uslubiyatning mohiyatini tashkil etadi". [B.B. Виноградов 1981:173-6.]. Demak, tildan foydalanish (qanday?) uslubiyat predmeti hisoblanadi.

Tilning asosiy vazifasi kommunikativ vazifadir. Lekin tildagi "...ayrim so'z va ibora (frazologizm)lar nutq egasining nutqi qaratilgan obyektga bo'lgan emotsional-ekspressiv munosabatini ifodalash xususiyatiga ega". [Sh.Shoaduraxmonov. 1980:130]. Shunga ko'ra, lug'at tarkibidagi so'zlar funksiyaviy uslubning muayyan turiga mansub, qo'llanishi mazkur uslub doirasida bo'ladi.

So'zlar qo'llanish doirasiga ko'ra: kitobiy, badiiy va oddiy so'zlashuvga xos guruhni tashkil etadi.

Kitobiy so'zlar, o'z navbatida, publitsistik, ilmiy va rasmiy ish yuritishga oid kichik guruhlariga ajratiladi: xabar, kub, bayonnoma.

Badiiy nutqqa xos so'zlar badiiylilik xususiyatiga ega bo'ladi: bo'y, balqimoq.

Kundalik oddiy va erkin muomalada qo'llaniladigan, adabiy nutq uchun deyarli me'yor sifatida qaralmaydigan so'zlar odatiy so'zlashuvga xos so'zlar hisoblanadi: ketvorgan, jinday, nuqul.

Ayni so'zlar kabi ibora (frazologizm)lar ham uslubning biror turiga moslashadi, lekin bunday iboralar kam uchraydi. Masalan, qurilish amalga oshadi, dalalarda ekinlar barq urib yashnaydi; senmisan, jonim qoqindiq.

Yuqoridagi misollar tarkibida kelgan iboralardan amalga oshmoq rasmiy (yozma) uslubga, barq urib yashnamoq badiiy uslubga, jonim qoqindiq oddiy so'zlashuv uslubiga xos.

Adabiy nutq tilshunoslik sohasida markaziy o'rinni egallaydi.

Hozirga o'zbek adabiy nutqidagi funktsiyaviy uslublarni quyidagicha tasniflash mumkin:
Kitobiy nutq (yozma):

- a) rasmiy ish yuritish uslubi; ilmiy uslub; publitsistik uslub (yozma);
- b) og'zaki publitsistik: kinonutq; telenutq.
2. Epistolyar (noma, xat, yozishma) uslub.
3. So'zlashuv nutqi (og'zaki, adabiy nutq).

Yuqoridagilar ikki katta guruhni tashkil etadi: yozma nutq; og'zaki nutq.

Matn tahriri jarayonida til va uslub birinchi bosqich hisoblanadi. Tahrirchining bu bosqichdagi vazifasi matn, avvalo, tasviriy tilshunoslik, so'ngra, adabiy nutq uslubiyati, ya'ni funksional uslubiyat talablariga mosligini ta'minlashdan iboratdir. Agar tahrirchi birinchi bosqich vazifasini bajarishga til va uslub har ikkisini qorishtirmay, ularni yaxshi farqlagan holda kirishsa yaxshi natijaga erishadi.

Foydalanilgan adabiyotlar:

1. Виноградов В.В. Проблемы русской стилистики. – Москва: 1981.
2. Шоабдурахмонов Ш., Асқарова М., Ҳожиёв А. ва б. Ҳозирги ўзбек адабий тили. – Т.: Ўқитувчи, 1980.
3. Taxirov Z.T. Adabiy tahrir. Darslik. –Т. Tafakkur-bo'stoni. Cho'lpon nomidagi NMIU, 2012.

IMPORTANCE OF STATISTICAL METHODS IN DETERMINE THE HOMONYMY

Axmedova Xolisxon Ilxomovna

PhD student

Tashkent state university Uzbek language
and literature named after Alisher Navai

Abstract: The problem of automatic processing of natural language remains relevant for more than half a century. One of the important problems in the field of NLP is the creation of a semantic analyzer, which in turn goes through a series of steps. Determining homonymy is important in the semantic analysis of sentences. Statistical methods are also used to determine homonymy. The frequency method is used to determine homonymy between grammatically similar word groups. This method involves extracting homonym classification parameters.

Keywords: NLP, semantic analyzer, homonymy, method based on statistical data, classification parameters, Hidden Markov model.

Аннотация: Проблема автоматической обработки естественного языка остается актуальной уже более полувека. Одной из важных задач в области НЛП является создание семантического анализатора, который в свою очередь проходит ряд этапов. Определение омонимии важно при смысловом анализе предложений. Статистические методы также используются для определения омонимии. Частотный метод используется для определения омонимии между грамматически сходными группами слов. Этот метод включает в себя извлечение параметров классификации омонимов.

Ключевые слова: NLP, семантический анализатор, омонимия, метод на основе статистических данных, параметры классификации, Скрытая марковская модель.

Natural Language Processing (NLP) – a general area of artificial intelligence and mathematicallinguistics, which studies the problems of computer analysis and synthesis of sentences in natural languages. Solving this problem means creating a more convenient form of interaction between a person and a computer. The problem of automatic processing of natural language remains relevant for more than half a century. The complexity of the problem and the lack of a clear idea indicate the difficulty of ways to solve it. Linguistic analyzers are particularly important as tools for automatic processing of sentences. Linguistic analyzers are divided into morphological, syntactic and semantic analyzers. They, are divided into several groups.

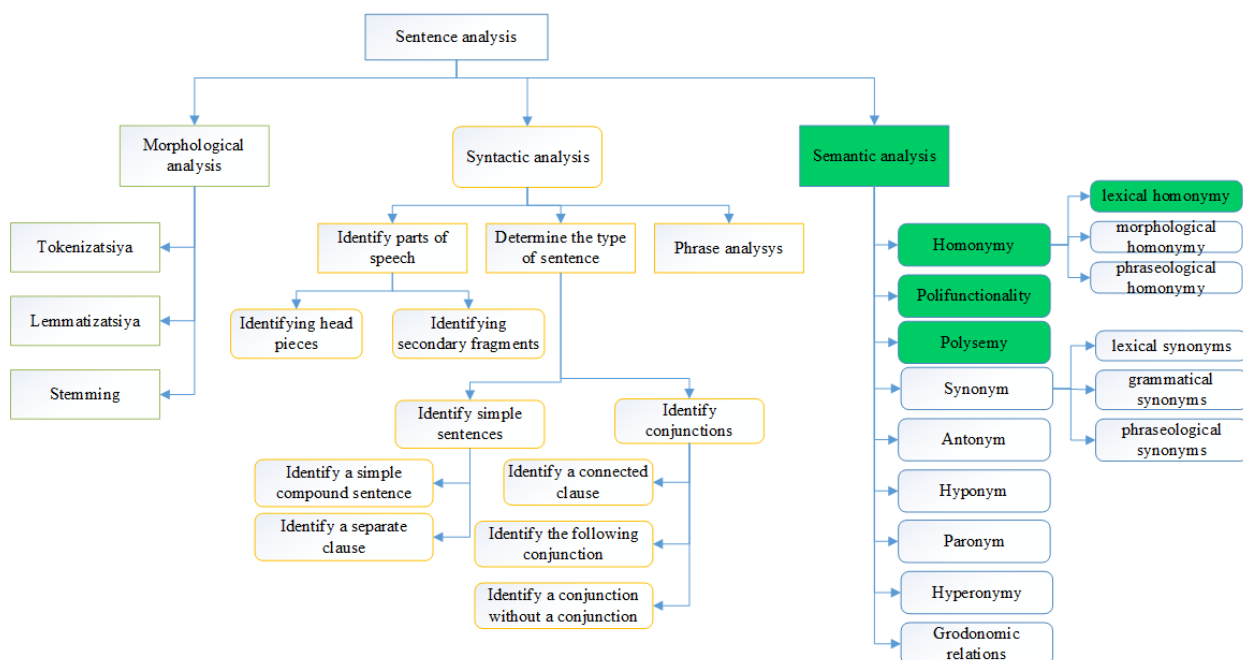


Figure 1. Hierarchy of linguistic analyzers and their elements

Each linguistic analyzer described in the hierarchy in Figure 1 has its own task. A morphological analyzer is a set of algorithms that compare individual words/word forms in the dictionary and determine the grammatical properties of words. A syntactic analyzer is an information system that syntactically analyzes an input sentence based on the characteristics of the input language.

A semantic analyzer is an information system that determines the content of the sentence through the meanings of the words in the sentence. Semantic search is performed through semantic analysis. The better the semantic analysis is developed, the more effective the search results will be. Implementation of semantic analysis directly depends on linguistic resources. Lexical resources include dictionaries, thesauruses and ontologies. Semantic analysis also has its own elements, which require a separate study. One of the important elements of semantic analysis is homonymy. Homonymy detection is interpreted differently in different natural languages. In world computer linguistics, 3 methods are mainly used in the semantic analysis of sentences:

- Rule-based methods;
- Method based on statistical data;
- A method based on machine learning.

These methods are used differently in different languages. Homonymy Baltabayeva J.K. and Sulaymanova J.N (Kazakh language) [2019], Ch.A. Davlyatova (Tajik language) [2017], V.V. Kukanova (Bashkir language) [2014], H. Heydarova (Azerbaijani language) [2017] V.V. Kukanova (Kalmyk language) [2011] has been the subject of research by Turkologist scholars. These Turkic scientists rely on the theory of homonymy developed by A.I. Smirnisky, V.V. Vinogradov, O.S. Akhmanova and other linguists. The results of the research of Turkic scientists show that the methods mentioned above are important for determining homonymy. In this article, we will try to highlight the use of statistical methods and its importance in distinguishing homonyms in the Uzbek language.

The method based on statistical data is used to solve the problem by classifying the grammatical parameters of words. These parameters are chosen differently in different natural languages. For example, when determining morphological homonymy in Russian, parameters such as word group, gender (genus) of the word, singular or plural form, lemma, lemma and word group, only lemma, and homonymy are distinguished [1]. The issue of removing morphological homonymy (snyatie homonymy) in the SemSin system of the Russian language is solved. In Russian, this term is called "snyatie homonym", because if the identified elements of the semantic analyzer are removed from the text, the analysis of the remaining elements increases the speed of the system, ensures the accurate and correct operation of the algorithms. Classification parameters for the Uzbek language can be divided as follows

Having studied foreign experiences in depth, we use rule-based, statistical data-based and machine learning-based methods to distinguish homonyms in the Uzbek language. When distinguishing homonyms in the Uzbek language, we divided them into groups such as homonyms within one word group, two word groups, three word groups, and four word groups according to their occurrence within word groups. We used the rule-based method to determine homonym within grammatically dissimilar word groups. This is covered in detail in articles [4,5,6].

The method based on statistical data} is used to solve the problem by classifying the grammatical parameters of words. These parameters are chosen differently in different natural languages. For example, when determining morphological homonym in Russian, parameters such as word group, gender (genus) of the word, singular or plural form, lemma, lemma and word group, only lemma, and homonym are distinguished [1]. The issue of removing morphological homonymy (snyatie omonimii) in the SemSin system of the Russian language is solved. In Russian, this term is called "snyatie omonimii", because if the identified elements of the semantic analyzer are removed from the text, the analysis of the remaining elements increases the speed of the system, ensures the accurate and correct operation of the algorithms. Classification parameters for the Uzbek language can be divided as follows in figure 2.

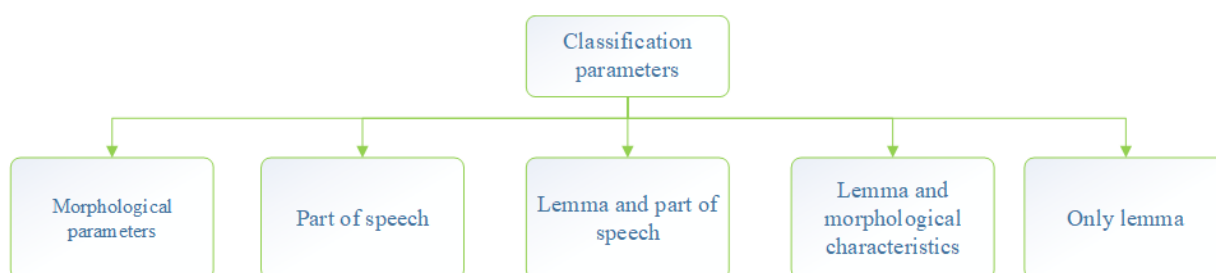


Figure 2. Classification parameters

If the level of accuracy of the semantic analysis of the word is not high, the classification parameters can be expanded. Statistical methods can be used using extracted parameters. The main task of this method is to divide the content of the context into n-grams, that is, the combinations of the input word in the context are determined and evaluated using evaluation methods [1]. In some statistical methods, a decision is made using the word and its suffixes, while in some, it is concluded using the semantic valence of the word in the context. It follows that methods based on statistical data are divided into several groups according to decision-making parameters (fig. 3).

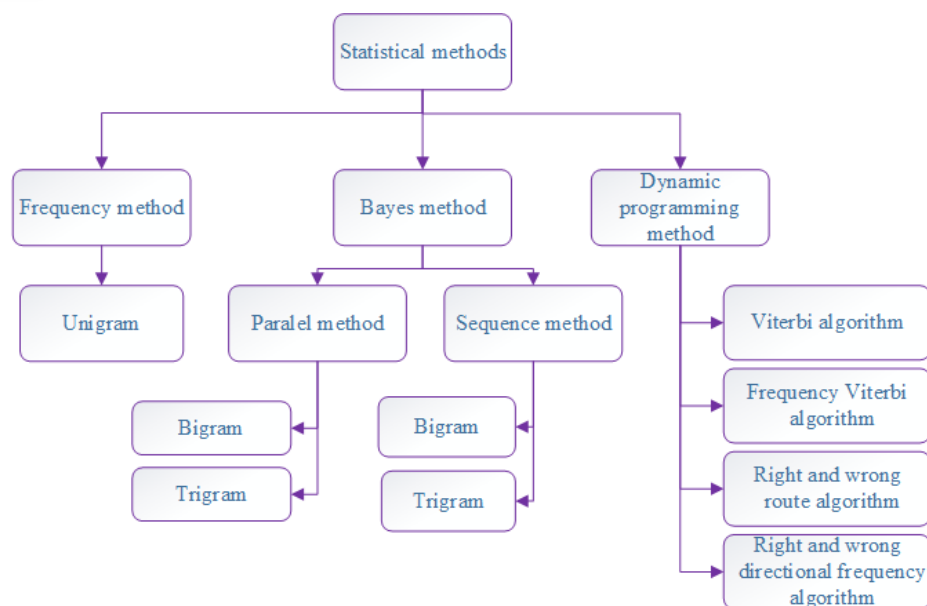


Figure 3. Hierarchy of statistical methods

As shown in Figure 1, the Frequentest method is semantically analyzed through the unigram, that is, the word itself, while the Bayesian method uses bigrams and trigrams of the word. Dynamic programming method uses n-grams. Context tokenization and tagging processes are carried out in the evaluation by these methods. Using statistical methods, we will consider the process of differentiating the meanings of homonyms within different word groups. For example, let's take a sequence of identifying homonym between noun or adjective word groups using the frequency method.

Homonyms within the *noun* $\vee$ *adjective* group are classified according to the following parameters

- Part of speech;
- Stem and lemma;
- Only lemma
- Only stem

Given a sentence with a homonym from the noun $\vee$ adjective group.

Isitma, yo'tal, burun bitishi va bosh og'rig'i, bularning hammasi gripp alomatlari hisoblanadi.

In this sentence, the word “alamat” is a homonym and has the following meanings.

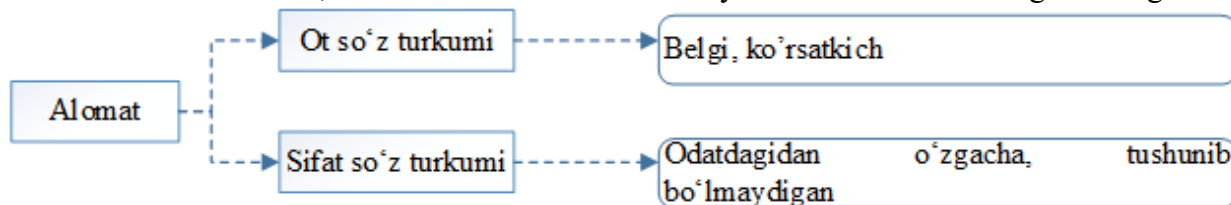


Figure-4. Definitions of alomat word

When classifying the word “alamat” according to the above parameters, we use the data of the national corpus of the Uzbek language. A total of 6,823 pieces of information were found

when the word “alamat” was searched through the Uzschoollcorpora.uz site, 100 of which were analyzed

79 of them are nouns

21 are adjectives

The analyzed 100 pieces of information were divided into core and appendices and the following results were determined.(fig.5)

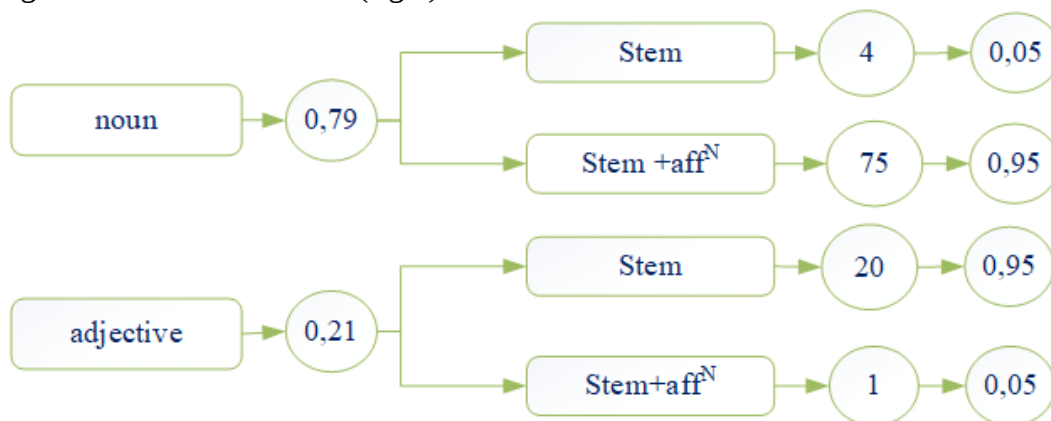


Figure 5.Statistical information

Based on the obtained statistical data, we also divide the homonym in the given sentence into stems and suffixes (fig.6)

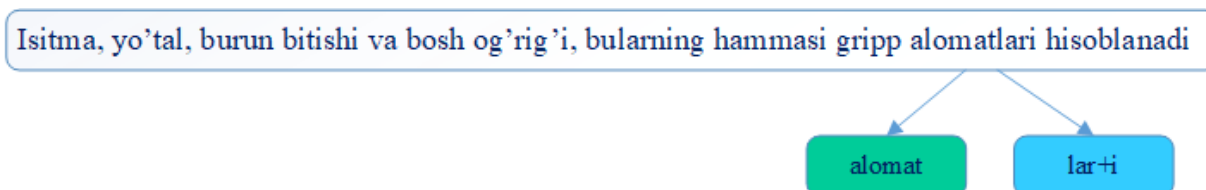


Figure 6. A sentence with the word “alamat”

So, based on the above statistics, the following decision is made for the word “alamat” in this sentence.(fig.7)

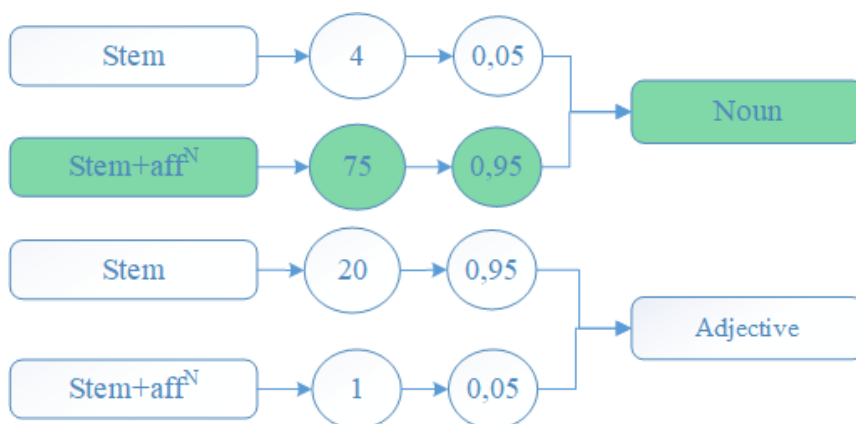


Figure 7. Homonym identification using classification parameters

The information in this chart shows that the word "alamat" in the given sentence is a homonym of the noun group with a 95% probability, and it means a sign, an indicator. In this way, homonym can be determined by frequency method. It can be seen that in order to use the frequency method, it is required to determine

the classification parameters for each of the homonyms within different word groups and perform statistical calculations based on them. Another common probabilistic approach is an algorithm based on the use of a Hidden Markov Model (HMM). The main idea of the algorithm is to choose a Grammar tag that maximizes the value of the following function for each word in the sentence:

$$P(\text{word} | \text{tag}) * P(\text{tag} | \text{previous } n \text{ tag})$$

Here, $P(\text{tag} | \text{previous } n \text{ tag})$ - conditional probability (estimated in the corpus), the probability of occurrence of the current tag with n predefined tags, $P(\text{word} | \text{tag})$ - conditional probability (calculated using corpus data) is a tag determined based on Grammatical properties of the word. Although HMM has complex computations, it has various simplifications in practice. Distinguishes word meanings with 96 % accuracy for English grammar. Applying this model to Russian may be difficult compared to English, requiring very large corpora given the richness of word formation and word variation in Russian.

In Tatar, as in other groups of agglutinative Turkic languages, morpheme is the most important meaningful unit of linguistics, carrying both semantic and syntactic information. TuganTel - an algorithm for determining morphological polysemy in the corpus of the Tatar language, a database has been developed.

Gataullin Ramil Raisovich discussed models and algorithms for removing homonym and polysemy in the Tatar language in his candidate thesis. The use of Markov models in distinguishing homonym and polysemy was also recommended in the Tatar language [2]. Having studied foreign experiences, it can be concluded that it is appropriate to use statistical methods to identify homonyms of different word groups in the Uzbek language.

References:

1. Rysakov S.V. Klyshinsky E.S.: Statisticheskie metody snyatiya homonym// Novye informatsionnye tehnologii v avtomatizirovannyx sistemax. – 2015 Springer, Heidelberg (2016). doi{10.10007/1234567890
2. Gataullin Ramil Raisovich.: Method, model and programmatic instrumental resolution of multivariateness in text// diss-VAK RF-05.13.11 kand.nauk.-(2019)
3. Chirag Goyal-June 23, 2021: <https://www.analyticsvidhya.com/blog/2021/06/part-9-step-by-step-guide-to-master-nlp-semantic-analysis/>
4. Elov B.B, Axmedova X.I.: Uchta so‘z turkumi doirasidagi omonimiyani farqlovchi biznes jarayonni modellashtirish//O‘zbekiston respublikasi innovatsion rivojlanish vazirligining, ilm-fan va innovasion rivojlanish ilmiy jurnal 2022 / 1
5. Sh.K. Gulyamova, X.I. Akhmedova. Linguistic basis of homonyms, mathematical model and algorithms (in the framework of nouns and verbs, adjectives and verbs) // KoqonDPI. Scientific reports, 20211`1
6. Axmedova X.I. O‘zbek tilidagi shakldosh so‘zlarni semantik tahlil qilish // “O‘zbekistonning umidli yoshlari” Mavzusidagi 7-son Republika ilmiy talabalar, magister yosh tadqiqotchilar va mustakil izlanuvchilar uchun onlayn anjuman sining materiallari heat. - Toshkent: tadqiqot, 2021.

O'ZBEK TILSHUNOSLIGI UCHUN QISQARTIRILMAYDIGAN GRAMMATIKA TUZISH.

Matlatipov G'ayrat Raximbayevich
Urganch davlat universiteti, t.f.n., dots

Otamuratov Maksudbek Shonazarovich,
Ramatova Ma'mura Gulimboy qizi
Urganch davlat universiteti 2-bosqich magistrantlari

Annotatsiya: Tilshunoslikda hisoblashni amalga oshirish uchun bir qancha modellar bo'lib, ular orqali ko'plab lingvistik muammolarni hal qilish mumkin. Fonologik, morfologik va sintaktik tahlillar, shuningdek, tokenizatsiya, so'z ma'nosini ajratib ko'rsatish, imloni tekshirish va tuzatish kabilarda sonli holat, grammatik transformatsiya, erishib bo'lmaydigan belgilarni olib tashlash va zanjir usullari, ayniqsa, samarali ekanligi isbotlangan sohalardir. Mazkur maqolada O'zbek tilshunosligi uchun KS-grammatikani normal shaklga keltirish orqali cheklangan miqdordagi shartli holatlarga ega bo'lgan har qanday tizim uchun matematik modelni tavsiflash holatlari tahlil qilingan. O'zbek tilshunosligi uchun KS-grammatikani normal shaklga keltirish amaliy misoli foydalanuvchi tomonidan ma'lum amalni bajarish orqali, tizim mos keladigan amallarni bajarish ko'rsatilgan.

Kalit so'zlar: KS-grammatika, zanjir qoida, kirish alfaviti, bo'sh qoida, steril, to'plamlar.

Annotation: In linguistics, there are several models for performing computation, through which many linguistic problems can be solved. Phonological, morphological and syntactic analyses, as well as tokenization, distinguishing word meaning, checking and correcting spelling are areas where numerical status, grammatical transformation, unattainable sign removal, and loop methods have been shown to be particularly effective. This article analyzes the cases of describing a mathematical model for any system with a finite number of conditional States by bringing KS-grammar into a normal form for Uzbek linguistics. For Uzbek linguistics, a practical example of bringing KS-grammar into normal form is shown by the user performing a certain action, performing the actions that the system corresponds to.

Keywords: KS-grammatics, chain rule, input alphabet, empty rule, sterile, sets.

Аннотация: В лингвистике существует несколько моделей реализации вычислений, с помощью которых можно решить многие лингвистические проблемы. Фонологический, морфологический и синтаксический анализы также являются областями, в которых числовое состояние, грамматическое преобразование, удаление недостижимых символов и методы цепочки оказались особенно эффективными в таких областях, как токенизация, выделение значений слов, проверка и исправление орфографии. В данной статье проанализированы случаи описания математической модели для любой системы с конечным числом условных состояний путем приведения КС-грамматики к нормальной для узбекской лингвистики форме. Практический пример приведения КС-грамматики к нормальной форме для узбекской лингвистики посредством выполнения пользователем определенного действия система показывает выполнение соответствующих действий.

Ключевые слова: KS-грамматика, цепное правило, вводный алфавит, пустое правило, стерильный, наборы.

KS grammatika qisqartirilgan grammatika deb yuritiladi. Grammatikalar orqali erishib bo'lmaydigan va yangi belgilar, sikllar va qoidalarini ("bo'sh" qoidalar) o'z ichiga olmaydigan grammatikalar qisqartirilmaydigan, ya'ni KS-grammatikadir. Biz oldin ko'rib chiqqan grammatikalar kanonik KS-grammatikalar deb ham ataladi

Ixtiyoriy KS-grammatikani qisqartirilgan shaklga aylantirish uchun biz quyidagi amallarni bajarib o'tishimiz zarur:

- Barcha bo'sh belgilarni olib tashlash;
- Barcha kirish mumkin bo'lmagan belgilarni olib tashlash;
- O'chirish - qoidalar;
- Zanjir qoidalarini olib tashlash.

Shuni ta'kidlash kerakki, konvertatsiya bosqichlari belgilangan tartibda bajarilishi kerak.

Grammatik transformatsiyalar. KS-grammatikasi til zanjirlarini yaratishda ishtirok etmaydigan va shuning uchun grammatikadan olib tashlanishi mumkin bo'lgan erishib bo'lmaydigan va yangi belgilarni o'z ichiga olishi mumkin. Buni quyidagi ta'riflar orqali ko'rib chiqishimiz mumkin.

Ta'rif: $A \in VN$ belgisi grammatikada **steril** deyiladi $G = (VT, VN, P, S)$ agar to'plam $\{ | \in VT^*, A \Rightarrow \}$ bo'sh.

Ta'rif: $x \in (VT \cup VN)$ belgisi, agar ushbu grammatika biron bir jumla shaklida bo'lmasa, $G = (VT, VN, P, S)$ grammatikasida erishib bo'lmaydigan deb ataladi.

Shundan ko'rinadiki, tugallanmagan belgilarni olib tashlash algoritmi uchun:

- Kirish: KS-grammatikasi $G = (VT, VN, P, S)$.
- Chiqish: KS-grammatikasi $G' = (VT, VN', P', S)$, tarkibida yangi belgilar mavjud emas, ular uchun $L(G) = L(G')$ bo'ladi.

Buning uchun quyidagi usuldan foydalanamiz:

N_0 , va N_1 , To'plamlarni rekursiv ravishda quramiz.

1. $N_0 = \{ \}$, $i = 1$.

2. $N_i = \{ A \mid (A >) \in P \text{ va } (A >) \in (N_{i-1} \cup VT)^* \} \cup N_{i-1}$.

3. Agar $N_i \neq N_{i-1}$, to $i=i+1$ va 2-bosqichga o'tamiz, aks holda $VN' = N_i$; P' faqat VN' VT dan belgilarni o'z ichiga olgan P to'plam qoidalaridan iborat;

$G' = (VT, VN', P', S)$ hosil bo'ladi.

Erishib bo'lmaydigan belgilarni olib tashlash algoritmi.

- Kirish: KS-grammatikasi $G = (VT, VN, P, S)$
- Chiqish: KS-grammatikasi $G' = (VT', VN', P', S)$ mavjud bo'lmagan belgilarni o'z ichiga olmaydi, ular uchun $L(G) = L(G')$

Buning uchun quyidagi usuldan foydalanamiz

1. $V_0 = \{ S \}$; $i = 1$.

2. $V_i = \{ x \mid x \in (VT \cup VN), (A > x) \in P \text{ va } A \in V_{i-1} \} \cup V_{i-1}$.

3. Agar $V_i \neq V_{i-1}$, to $i=i+1$ va 2-bosqichga o'tamiz, aks holda $VN' = V_i$; $VT' = V_i$; P' faqat V_i dan belgilarni o'z ichiga olgan P to'plam qoidalaridan iborat; $G' = (VT', VN', P', S)$ tuzib olamiz. Belgilarni o'chirish ushbu belgilarni o'z ichiga olgan xulosa qoidalarini o'chirish bilan birga keladi.

Agar ushbu algoritmda (1) va (2) bosqichlar qayta tartibga solinsa, natija har doim ham to'g'ri bo'lmaydi.

Misol tariqasida, $S > U$, $S > VZ$, $T > aa$, $T > bb$, $U > aUa$, $U > bUb$, $V > aTb$, $V > bTa$, $W > YZY$, $W > aab$, $X > Xa$, $X > Xb$, $X > e$, $Y > YY$, $Y > aU$, $Y > e$, $Z > W$, $Z > b$. qoidalariga ega kontekstsiz G grammatikasini ko'rib chiqamiz. U bepust belgisini o'z ichiga olgan to'rtta

qoidani olib tashlasak, biz G_1 grammatikasini olamiz: $S > VZ, T > aa, T > bb, V > aTb, V > bTa, W > YZY, W > aab, X > Xa, X > Xb, X > e, Y > YY, Y > e, Z > W, Z > b$. Unda X belgisiga erishib bo'lmaydi. X o'z ichiga olgan uchta qoidani olib tashlasak, biz qoidalar bilan G_2 grammatikasini olamiz: $S > VZ, T > aa, T > bb, V > aTb, V > bTa, W > YZY, W > aab, Y > YY, Y > e, Z > V, Z > b$. Shubhasiz, $L(G) = L(G_2)$ va grammatika G_2 yangi va erishib bo'lmaydigan belgilarni o'z ichiga olmaydi.

Qisqartirilmagan grammatikalarni o'zgartirish uchun ko'rib chiqilayotgan o'zgarishlarning oxirgi turi grammatikadan bo'sh o'ng tomoni bo'lgan qoidalarni olib tashlash bilan bog'liq.

Ta'rif. $A >$ ko'rinishdagi qoida "bo'sh" (bekor qiluvchi) qoida deyiladi.

Ta'rif. Grammatika qisqartirilmaydigan yoki "bo'sh" qoidalari bo'lmagan grammatika deyiladi. Agar:

- 1) Grammatika sxemasi bekor qilish qoidalarini o'z ichiga olmaydi;
- 2) Yoki grammatika sxemasi $S >$ shaklining faqat bitta qoidasini o'z ichiga oladi, bu yerda S grammatikaning boshlang'ich belgisidir va S belgisi qolgan grammatik qoidalarining o'ng qismlarida uchramaydi.

Yo'q qilish qoidalarini o'z ichiga olgan grammatikalar uchun quyidagi tasdiq to'g'ri.

Ma'lumot. Tasdiqlangan qoidalarini o'z ichiga olgan har bir KS-grammatikasi G' uchun $L(G')=L(G)$ ga teng qisqartirilmaydigan G grammatikasini qurish mumkin. Qisqartirmaydigan grammatikani qurish, yo'q qiluvchi qoidalarining terminal bo'lmaganlarini yo'q qilish natijasida qo'shimcha qoidalarni qurish hisobiga berilgan grammatika qoidalarining ko'payishiga olib keladi. Qo'shimcha qoidalarni yaratish uchun barcha grammatik qoidalaridagi noterminal o'rniga bo'sh qatorni barcha mumkin bo'lgan almashtirishlarni bajarish kerak.

Agar grammatikada $S >$ shaklidagi qoida mavjud bo'lsa, bu yerda S grammatikaning boshlang'ich belgisi bo'lib, S belgisi boshqa grammatik qoidalarining o'ng qismlariga kiritilgan bo'lsa, unda siz yangi S' boshlang'ich belgisini kiritib, o'rniga S ni kiritishingiz kerak. $S >$ qoidasi ikkita yangi qoida bilan: $S' >$ va $S' > S$. Qisqartirilmaydigan grammatikalarining tuzilishiga misol sifatida keling, quyidagi grammatikadan bekor qilish qoidalarini olib tashlaylik:

$G(\{a, b\}, \{S\}, P = \{S > aSbS, S > bSaS, S > \}, S)$.

Grammatikaning **birinchi qoidasida** S belgisini barcha mumkin bo'lgan almashtirishlarni bajarib, biz shaklning to'rtta qoidasini olamiz:

$S > aSbS, S > abS, S > aSb, S > ab$.

Ikkinchi qoidaga o'xshab, bizda:

$S > bSaS, S > baS, S > bSa, S > ba$.

Yo'q qiluvchi qoidani tashkil etuvchi boshlang'ich belgi boshqa grammatik qoidalarining o'ng qismlariga kiritilganligini hisobga olib, $S >$ qoidasini $S' >$ va $S' > S$ ko'rinishdagi qoidalar bilan almashtiramiz. Tuzilgan qoidalar to'plami istalgan qisqartirilmaydigan grammatikaning qoidalari to'plamini tashkil qiladi.

$S' > S, S > aSbS | abs | aSb | ab | bSaS | ?baS | bSa | ba$

Grammatikaning yuqoridagi barcha o'zgarishlari ham chekli, ham surish avtomatlarini qurishda qo'llanilishi mumkin.

Keling, endi kutilmaganda kelib chiqadigan hodisalarni ko'rib chiqamiz. **Ta'rif.** $A > B$ shaklidagi grammatika qoidasi, bunda $A, B \in V_N$, zanjir qoidasi deyiladi.

Zanjir qoidalarini o'z ichiga olgan KS-grammatikasi G uchun zanjir qoidalariga ega bo'lmagan ekvivalent G' grammatikasini qurish mumkin

Ta'rifning mazmuni quyidagicha. Agar G grammatikasi $A > B, B > C, C > aX$ qoidalariga ega bo'lsa, bunday qoidalarni bitta $A > aX$ qoidasi bilan almashtirish mumkin, chunki aX

zanjirining $A \Rightarrow B \Rightarrow C \Rightarrow aX$ hosilasi, grammatika G' grammatikasida $A > aX$ qoidasi yordamida olinishi mumkin.

1. Har bir $A \in N$ konstruksiya uchun $N_A = \{B | A \Rightarrow^* B\}$ quyidagicha:

a) $N_0 = \{A\}$ va $i=1$ ni qo'ying.

b) $N_i = \{C | B > C \text{ to'plami } P \text{ va } B \in N_{i-1}\} \cup N_{i-1}$ ga tegishli.

c) Agar $N_i \neq \emptyset$ bo'lsa N_{i+1} , $i = i+1$ qo'ying va (b) qadamni takrorlang. Aks holda $N_A = N_i$ qo'ying.

2. P' ni quyidagicha tuzing: agar $B > b$ P ga tegishli bo'lsa va zanjir qoidasi bo'lmasa, P' ga $B \in N_A$ bo'ladigan barcha bunday A uchun $A > b$ qoidasini kiriting.

3. $G' = (VT, VN, P', S)$ qo'ying.

Misol tariqasida, biz G grammatikasidan zanjirli qoidalarni olib tashlashni amalga oshiramiz:

$G = (\{+, *, (,), a\}, \{E, T, F\}, P = \{E > E+T | T, T > T * F | F, F > (E) | a\}, E)$.

Keling, grammatik qoidalarni ikkita kichik to'plamga ajratamiz:

$P_1 = \{E > T, T > F\},$

$P_2 = \{E > E+T, T > T * F, F > (E) | a\}$

(1) bosqichda $N_E = \{E, T, F\}$, $N_T = \{T, F\}$, $N_F = \{F\}$. (2) qadamdan so'ng P' to'plamiga aylanadi.

$E > T+T | T * F | (E) | a$

$T > T * F | (E) | a$

$F > (E) |$

Shu tariqa, o'zbek tilshunosligi uchun qisqartirilmaydigan grammatikalarni tuzishimiz mumkin.

Foydalanilgan adabiyotlar:

1. Ben-Ari, M., Mondada, F. (2018). Finite State Machines. In: Elements of Robotics. Springer, Cham. https://doi.org/10.1007/978-3-319-62533-1_4
2. Wang, Jiacun (2019). Formal Methods in Computer Science. CRC Press. p. 34. ISBN 978-1-4987-7532-8.
3. Nelson, R. Context-Free Grammars. Retrieved June 11, 2016, from https://www.cs.rochester.edu/~nelson/courses/csc_173/grammars/cfg.html
4. Belzer, Jack; Holzman, Albert George; Kent, Allen (1975). Encyclopedia of Computer Science and Technology. Vol. 25. USA: CRC Press. p. 73. ISBN 978-0-8247-2275-3.

O'ZBEK TILI MATNIDAGI NOSTANDART SO'ZLARNI NORMALLASHTIRISH MUAMMOLARINI YECHISH

Boburxon Shuhrat o'g'li

Muhammad al-Xorazmiy nomidagi TATU tayanch doktoranti

Annotatsiya: Matnni normallashtirish text-to-speech tizimining muhim tarkibiy qismi bo'lib, matnni normallashtirishdagi qiyinchilik nostandart so'zlarni (non-standart words) ajratishdir. Ushbu maqola keng miqyosli o'zbek tili nutqi asosida nostandart so'zlarning taksonomiyasini ishlab chiqadi va ikki bosqichli nostandart so'zlarni aniqlashtirish strategiyasini taklif qiladi, dastlabki tasniflash uchun cheklangan davlat avtomatlari va pastki sinflarni aniqlash uchun maksimal entropiya tasniflagichlari. Yuqoridagi nostandart so'zlar taksonomiyasiga asoslanib, ikki bosqichli yondashuv ochiq testda 98,53% aniqlikka erishadi. Tajribalar shuni ko'rsatadiki, nostandart so'zlar taksonomiyasining yuqori boshlang'ich ishlashini ta'minlaydi va maksimal entropiya tasniflagichlari sezilarli yaxshilanishlarga erishadi.

Matnni normallashtirish TTS tizimlarida matn tahlilining hal qiluvchi komponentidir. O'zbek tili matnida juda ko'p nostandart so'zlar mavjud, chunki ularning xususiyatlarini lug'atda topib bo'lmaydi va «harfdan tovushga» qoidalarini qo'llash orqali ularning talaffuzini topib bo'lmaydi [1]. Nostandart so'zlarni mos keladigan standart so'zlarga normallashtirish kerak va bunday jarayon matnni normallashtirish deb ataladi. Ingliz tilida raqam ifodalari va qisqartmalar nostandart so'zlar hisoblanadi. Hatto jummalarni segmentatsiyalash ham matnni normallashtirish vazifasiga kiradi. O'zbek tili uchun raqamlar, belgilar va alifbolar kabi o'zbekcha bo'lmagan so'zlarni o'zbekcha shakllarga normallashtirish kerak. Nostandart so'zlar mahalliy kontekstga va matn janriga qarab turli xil standart so'zlarga o'zgartirilishi mumkin. Demak, umuman olganda bu, o'ta og'ir omografni aniqlash vazifasidir [2]. Nuance Vocalizer-da asosiy dastur kodining 20% dan ortig'i (kod metrikasi qatori) matnni normallashtirishga bag'ishlangan va yangi kiritish shakllari qo'shilishda davom etmoqda [3]. Matnni normallashtirishning odatiy usullari oddiy tayyorlangan qoidalarga asoslanadi. Ammo bunday oddiy tayyorlangan qoidalarini yozish, saqlash va yangi domenlarga moslashish qiyin. Boshqa tomondan, gomografni aniqlashni hisobga olgan holda, mashinali o'qitishning ko'plab usullari qo'llaniladi va ularning afzalliklarini ko'rsatdi. Qarorlar daraxti va qarorlar ro'yxati ingliz va hind tilidagi matnni normallashtirishda qo'llaniladi, xususan o'zbek tilida ham [4]. Support vektor machine (SVM) klassifikatsiya algoritmi tasnifiga ko'ra matn ma'lumotlari sinflashtiriladi va qo'llaniladi [5].

Biroq o'zbekcha matnni normallashtirish modullarining aksariyati qoidalarga asoslangan va so'zlarni segmentatsiyalash jarayonidan oldin amalga oshiriladi. Chunki o'zbekcha matnda so'zlar orasida bo'sh joylar har xil holatda amalga oshiriladi. Adabiyotda tashqi qoidaga asoslangan o'zbekcha matnni normallashtirish usulini qabul qiladi. U 15 dan ortiq tashqi qoidalarini hamda so'z va nutq ma'lumotlaridan foydalanadi. Yana boshqalar so'z segmentatsiyasini, nomli obyektini tanib olish va nostandart so'zlar jarayonini birlashtirilgan ramkaga qo'ydi.

Ushbu maqolada taklif qilingan matnni normallashtirish usuli so'zlarni segmentatsiyalash jarayonini talab qilmaydi. Cheklangan holat avtomatlari matndan nostandart so'zlarni aniqlaydi va

dastlabki tasnifni amalga oshiradi, so'ngra keyingi tasniflash uchun maksimal entropiya tasniflagichlaridan foydalaniladi.

Nostandart so'z taksonomiyasi keng miqyosli TTS korpus tizimli tekshiruvdan so'ng ishlab chiqilgan. Ushbu taksonomiya asosida uch qatlamli normallashtirish jarayoni ishlab chiqilgan. Cheklangan holat avtomatlari nostandart so'zlarni aniqlash va dastlabki tasniflash uchun ishlatiladi. Nostandart so'zlarni keyingi tasniflash uchun maksimal entropiya tasniflagichlari qo'llaniladi va sonli holat transduserlari standart so'zlarni yaratish uchun ishlatiladi.

Nostandart so'z taksonomiyasi matnni normallashtirishning asosidir. U nostandart so'z toifalarini belgilaydi, ularga ko'ra nostandart so'zlar aniqlanadi, tasniflanadi va o'zgartiriladi. O'zbek tili matnida arab, rim raqamlari va ba'zi belgilar normallashtiriladigan asosiy obyektlardir. Taksonomiya 30 raqamli qatorni o'z ichiga olgan.

Jadval 1. Kirish formatlariga asoslangan nostandart so'zlar taksonomiyasi

Raqamlar	raqamlar	1,2,3, va hokazo...
	nuqta	1.29, 2000.9.10, 162.105.81.14, ...
	defis	1998-2002, 2000-9-10, 4-3-2-1, ...
	slesh	1/3, 2000/9/10, ...
	ko'rsatgich	10:15, 10:15:20, ...
	qo'shimchalar	%,s(o'n ming), sifatlovchilar, ...
	diapazoni	100-200 D(100 dan 200 kishigacha), ...
	boshqalar	'99, ...
Simvollar	-, /, :, ., ×, >, =,	
Boshqalar	URL, Email, Alfabetlar, ...	

1-jadval nostandart so'z taksonomiyasining qisqacha tavsifi keltirilgan. Nostandart so'zlar birinchi navbatda formatlari bo'yicha tasniflanadi. Algoritmdagi 276 nostandart so'zning 95% raqamlar qatorlari, raqamli qatorlar va belgilarning turli kombinatsiyalari (nuqta, defis, slash, ikki nuqta va boshqalar) va o'zbekcha kvalifikatsiyalar kabi qo'shimchali raqam satrlarini o'z ichiga olgan son ifodalari. Belgilar o'zgartirilishi kerak bo'lgan yana bir toifadir va ba'zi belgilar ko'p talaffuzga ega. URL va elektron pochta manzillarini normallashtirish aniq. Ingliz alifbosi satrlari o'zlarining tegishli o'zbekcha tarjimalariga ega. Boshqa barcha noyob nostandart so'zlar ham "Boshqalar" toifasiga qo'shiladi. Taksonomiyaga jami 48 turdagi turli formatdagi nostandart so'zlar kiritilgan. Bu turlarning ba'zilar aniq talaffuzga ega, boshqalari esa yo'q.

Talaffuzi formatlar bilan belgilanadigan nostandart so'zlar Asosiy nostandart so'zlar (Bnostandart so'zs), noaniq nostandart so'zlar esa noaniq nostandart so'z (Anostandart so'z) deb nomlanadi. 2-jadval va 3-jadval mos ravishda Bnostandart so'z va Anostandart so'zlarga ba'zi misollar keltiradi. 2-jadvalda Bnostandart so'zlarning algoritmda taqsimlanishi ko'rsatilgan. Ko'rinib turibdiki, Bnostandart so'z barcha nostandart so'z hodisalarining taxminan 84% ni tashkil qiladi va "miqdor" barcha bu hodisalarining 55% ni tashkil qiladi. Bu shuni anglatadiki, nostandart so'zlarning 84% talaffuzi ularning formatlari bilan belgilanadi va faqat 16% noaniqdir. "Miqdorda" o'zbekcha sifatlovchilar va o'lchovlar kabi qo'shimchalar nostandart so'z talaffuzlarini aniqlash uchun muhim vositadir.

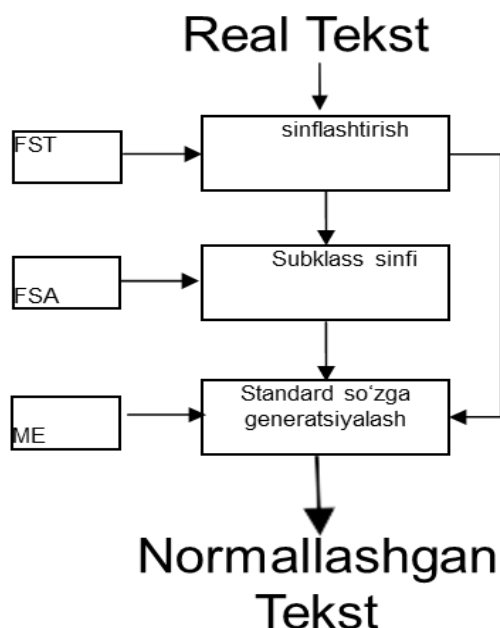
Jadval 2. Bnostandart soʻzlarga misollari

Nostandart soʻzlar sinfi	Misol	Foiz
Koʻrsatgichlar	35 P inchi,nchi	55%
Butunlar	100 \$	8%
Foizlar	10%, 12.5%	6%
Sana	27 oktabr	4%
Raqam va soʻzlar	15 ming	3%
Raqamlar bazisi	5 kg, 10 cm	2%
Yillar	5 yil	2%
Boshqalar	Win32	4%

Jadval 3. Anostandart soʻzlarga misollari

Nostandart soʻzlar sinfi	Soʻzlar	misol
Raqamlar	Raqamga raqam	2 ga 11 (2.11 metr)
	butun	110
	Ovoz bersh	110
	Inliz alifbosi	p2p
a-giper	yil-yil	1998-1999
	telefon	+99893 385 34 34
	raqam-raqam	737-200 (Boying737-200)
	son-son	200-300
	rate	2-3
	Subtract	100-1=99
Slesh	Fraction	1/3
	sana	2001/01
Dominant	Vaqt	10:15 (10:15 soat)
	Qadam	10:15

3-jadvalda javoblarning ayrim toifalari va ularni talaffuz qilishning mumkin boʻlgan usullari koʻrsatilgan. Koʻrinib turibdiki, baʼzi nostandart soʻzlar yuqori darajada noaniqliklarga ega va ularning maʼnosi ichki va kontekstli maʼlumotlarga muhtoj. Yuqoridagi taksonomiyaga asoslanib, matnni normallashtirish jarayoni uch bosqichdan iborat. Birinchi bosqichda haqiqiy matndan nostandart soʻzlarni aniqlash va dastlabki tasniflash uchun mashinali oʻqitish algoritmdan foydalaniladi. Ushbu bosqichda Bnostandart soʻz tasnifi tugallanadi. Javob berish uchun boshlangʻich tasnifning chiqishi subklassni aniqlash uchun ishlatiladi. Maksimal entropiya klassifikatorlari pastki sinfni aniqlashtirish modulida qoʻllaniladi. Nostandart soʻzlar sinf yorligʻi bilan belgilangan boʻlsa, Cheklangan holat oʻtkazgich uni standart soʻzlarga aylantiradi. Jarayon oqimi 1-rasmda tasvirlangan va uning ishlash jarayoni 2,3 - rasmda koʻrsatib oʻtilgan.



1-rasm. Matnni normallashtirish sxemasi

o'zbek tilidagi ixtiyoriy matn 	Normallashtirish
--	--------------------------------------

2 – rasm. tizim interfeysi

XI - asr 2001-yil 10 - noyabr 	o'n birinchi asr ikki ming birinchi yil o'ninchi noyabr
---	---

3 – rasm. Dasturdan foydalanish natijasi

Xulosa. Ushbu maqola o'zbekcha matnni normallashtirish bo'yicha keng qamrovli tekshiruv o'tkazadi. Nostandard so'z taksonomiyasi keng miqyosli korpus asosida ishlab chiqilgan. Taksonomiyani tizimli tahlil qilgandan so'ng, ikki bosqichli nostandard so'zlarni tasniflash strategiyasi, dastlabki tasniflash uchun chekli holat avtomatlari va keyingi tasniflash uchun maksimal entropiya tasniflagichlari taklif etiladi. Tajriba natijalari shuni ko'rsatadiki, bu yondashuv yaxshi samaradorlikka erishadi va yangi domenlar uchun yaxshi umumlashtiriladi. Bundan tashqari, bu yondashuv xarakterga asoslangan bo'lib, so'zlarni segmentatsiyalash jarayonini talab qilmaydi.

Foydalanilgan adabiyotlar:

1. Richard Sproat, Alan Black, Stanley Chen, Shankar Kumar, Marsi Ostendorf, and Christopher Richards, "Normalization of Non-Standard Words," *Computer Speech and Language*, 15(3):pp.287-333, 2001.
2. 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;5(03):89-100.
3. Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aforef." (2018).
 4. Musaev M. M., «Sovremennyye metody sifrovoy obrabotki rechevyykh signalov.,» Vestnik TUIT, t. 2, № 42, pp. str. 2-13, 2017[In Russian].
 5. Musaev M.M., Xujayarov I.Sh., Ochilov M.M., «Mashinali o‘qitish algoritmlari asosida o‘zbek tili fonemalarini tanib olish.,» Informatika va energetika muammolari., t. 6, 2019[In Uzbek].
 6. Alimuradov A.K.,Churakov P.P., «Obzor i klassifikatsiya metodov obrabotki rechevyykh signalov v sistemax raspoznavaniya rechi» Izmerenie. Monitoring. Upravlenie. Kontrol, t. №2, № 12, pp. 27-35, 2015[In Russian].
 7. Musaev M., Khujayorov I. and Ochilov M., «The Use of Neural Networks to Improve the Recognition Accuracy of Explosive and Unvoiced Phonemes in Uzbek Language» v Information Communication Technologies Conference (ICTC), Nanjing, China, 2020.
 8. Musaev M. M., Raximov M. F. «Algoritmy parallelnoy obrabotki rechevyykh signalov.,» Vestnik TUIT, t. 2, № 46, pp. 2-13 str., 2018[In Russian].
 9. M.M. Musaev, U.A. Berdanov, K.E. Shukurov, «Hardware and software solution signal compression algorithms based on the Chebyshev polynomial» International Journal of Information and Electronics Engineering, t. Vol. 4, № No. 5, pp. pp. 380-383, 2014.

MASHINA TARJIMASIDA ANTROPONIMLAR BILAN ISHLASH (O'ZBEK VA TURK TILLARIDAGI ANTROPONIMLAR MISOLIDA)

Avezova Anajon Ko'paysin qizi

O'zbekiston Milliy universiteti

O'zbek filologiyasi fakulteti

2-kurs magistranti

E-mail: Avezovaanajon9309@mail.ru

Annotatsiya: Ushbu maqolada mashina tarjimasida antroponimlar masalalari ko'rib chiqiladi va mashina tarjimasida atoqli nomlarga yondashuvlar tavsiflanadi, shuningdek, tegishli nomlarni tarjima qilishda yuzaga keladigan muammolar va ilmiy adabiyotlarda taklif qilingan yechimlar ko'rib chiqiladi. O'zbek va turk tillaridagi antroponimlar misolida muammo yoritiladi.

Kalit so'zlar: antroponimlar, mashina tarjimasi, atoqli otlar, tarjima, matn, til texnologiyasi.

Annotation: The issues of the transfer of proper names in machine translation are considered, approaches to the transfer of proper names are described, as well as problems arising in the translation of proper names, and solutions that are proposed in the scientific literature. The problem is illustrated by the example of anthroponyms in the Uzbek and Turkish languages.

Keywords: anthroponyms, machine translation, proper names, translation, text, language technology.

Аннотация: Рассматриваются вопросы передачи имён собственных в машинном переводе, описаны подходы к передаче имён собственных, а также проблемы, возникающие при переводе имён собственных, и решения, которые предложены в научной литературе. Проблема объясняется на примере антропонимов в узбекском и турецком языках.

Ключевые слова: машинный перевод, имёна собственные, перевод, текст, языковая технология.

O'zbek tilidagi atoqli otlar mashina tarjimasida mumoli masaladir. O'zbek tilida hech qanday tarjimaga muhtoj bo'lmagan atoqli nomlar guruhi mavjud. Shaxs nomlari bunga misol bo'la oladi. Shuningdek, tarjima qilinishi kerak bo'lgan atoqli nomlar ham bor. Bunday nomlarga bosh harflar bilan yozilgan tashkilot nomlari (masalan, Birlashgan Millatlar Tashkiloti), kitob, gazeta, kinofilm nomlari, tarixiy voqealar va bayram nomlari misol bo'ladi.

Atoqli nomlarni tarjima qilish muammosiga to'xtaladigan bo'lsak, "onomastika" atamasiga murojaat qilish kerak. Onomastika (boshqa yunoncha onomostukdan) - ism berish san'ati - leksikologiyaning tegishli nomlarni o'rganadigan bo'limi [Горкин, 2006; 600]. D.I.Ermolovichning fikricha, "Atoqli otlar tasvirlangan vaziyatdan qat'iy nazar va majburiy holda obyektini maxsus, individual belgilash uchun xizmat qiladi. Umuman olganda, nomlar avtomatik ravishda tarjima qilinadi. Biroq bu yondashuv matnlarni tarjima qilishda, boshqa tillardan ism va unvonlardan foydalanishda xato va noaniqliklarga olib keladi. Ushbu muammo yetarli darajada tadqiqot olib borilmaganligi sababli yuzaga keladi. [Ермолович, 2001; 6]

Birinchi navbatda ko'rinadigan yechim barcha atoqli nomlarni tahlil tizimiga yoki morfologik analizatorga yoki mashina tarjimasi dasturining keyingi ishlov berish moduliga

kiritishdir. Ammo amalda bunday ro'yxat har doim to'liq bo'lmaydi, chunki yangi onimlar doimiy ravishda paydo bo'ladi.

Til texnologiyasi va tilshunosligi bo'yicha Finlyandiya olimi Arvi Yoxannes Xurskainen faqat uzatilishi kerak bo'lgan atoqli nomlarni, ya'ni o'z so'z shakliga qarab aniqlab bo'lmaydigan onimlarni sanab o'tishning eng yaxshi usuli deb hisoblaydi. Gapning boshlang'ich so'zlari bo'lmagan, lekin bosh harf bilan yozilgan qolgan so'zlar dastur tomonidan o'zinkidek talqin qilinadi [Hurskainen, 2013;1-5].

Tabiat va osmon jismlari bilan bog'liq ismlar

Turkiy xalqlarda ba'zi tabiat hodisalari (Yomg'ir, Chaqmoq, Yashin va hokazo) va bola tug'ilgan kuni sodir bo'lgan muhim voqeaga ham ism sifatida qo'yilishi mumkin. Masalan, orolda tug'ilgan bolaga Orol nomini berish, kemada tug'ilgan bolaga "Deniz", deb nom berish odatiy holga aylandi. Turkiyada Gökçe, Gökçen, Gündeş, Seher, Şafak, Leyla" bolalar uchun ismlar sifatida berilgan. Bulardan tashqari turk oilalari bolalarga fasl nomlari (Bahor, Yoz atirgul, Kuz, Hazan va hokazo) va quyosh taqvimidagi oy nomlarini (Nisan (aprel), Eylul (sentyabr), Haziran (iyun) va h.k.) berganlar. Bundan tashqari, O'g'uz turklarida oyning 1-kunida tug'ilgan qizlarga Aybike (oy xonimi) deyiladi.

Turkiyada oilalar farzandlariga mevalarning nomini ham qo'yishadi 23 ta lug'atda 109 xil meva nomlari mavjud. Bu ismlar orasida asosan qizlarning ismlariga afzallik beriladi. O'simlik va mevalardan ilhomlangan 9 ta erkak ismlari kuzatilgan bo'lsa, bu raqam ayollarda 39 ta kuzatildi. Umumiy aholi orasida har 100 000 erkakdan 337 tasi va har 100 000 ayoldan 1256 tasi o'simliklar va mevalardan ilhomlangan nomga ega. Bolalarning mevalarning nomlari bilan nomlanishining sababi bolalarda mevalarning rangi, hidi, ta'mi va shakli kabi go'zal xususiyatlarini ko'rish istagi, bu nomlarning aytilishi va nomlarning ma'lum davrlarda qo'llanilishi sifatida ko'rsatish mumkin. Bolalarga beriladigan mevalarning nomlari: Malina (Ahududu) Oltinbosh (Altınbaş), Badem, Baran, Bostan, Böğürtlen, Çağla, Çakır, Çilek, Erik, Fındık, Gökçe, Hurma, İdris, İğde, Kiraz, Limon, Melike, Nar, Pelit, Portakal, Salkım, Semra, Şekerpare

Gul nomlari va boshqa o'simliklar bolalar uchun nom sifatida afzal bo'lishi mumkin. Çam, karaağaç, kavak, meşe ve ardıçla ilgili Narat, Karagaç, Karagay, Tirek, Biktirek, Baytirek, Büztirek, İmen, İmenbay, İmengol, İmenkildi, İmenkiske, Artış

Turk jamiyatlarida oilalar bolaning tug'ilgan kuni, fasli va yiliga qarab ism qo'yishlari mumkin. Ob-havo sharoiti nomga ham ta'sir qilishi mumkin. O'rim-yig'im davrida O'roqboy, Urakay, Urakbay, Urakbike, Urakchi kabi nomlar, Saban Shudgor mavsumida Sabanay, Sabanak, Sabanchi kabi nomlar ism sifatida berilgan bo'lsa, Urmanay, Urmanbay, Urmanchi ismlari qisqartirilgan, Qo'zibala, Qo'zibek, Qo'ziqildi, Qo'zimko'l kabi ismlar tug'ilgan bolaning ismi sifatida berilgan. Shuningdek, xirmon oyi va Indirchi nomi berilgan. Kunning ma'lum bir vaqtida tug'ilgan bolalarga Tanatar, Tanbike, Tangatar, Tanatmish, Tanchi, Tankildi, Tongatar ismlari qo'yilgan. Bayram va maxsus kunlarda tug'ilgan bolalarga Beyrem, Beyrembay, Beyrembike, Beyremhan, Beyremsilu, Bayraş, Kileş, Chuk, Chukhan, Kankay, Baganay, Tinibek kabi ismlar qo'yilgan.

12 ta hayvondan iborat turk taqvimidagi hayvon nomi, har oyda har xil hayvon nomi ham shaxs ismlari sifatida ishlatiladi.

Turk oilalari farzandlariga tug'ilish tartibiga ko'ra ism qo'yishlari ham ko'rinib turibdi. Xususan, birinchi va oxirgi bolaning ismlari ularning tug'ilish tartibidagi o'rnini aniqlashda samarali. Birinchi Farzandga Ilknur, Ilkay, Ilker, Ilksen, Adam; Sani ikkinchi bolaga; oxirgi

bolaning ismi Kenja, Kenjaboy, To'xtasin, Songül, Soner, Dursun, Sonay, Yeter. Bu ismni qabul qilganimizdan keyin tez-tez uchrab turadigan yana bir nomlash an'anasidir.

Uzoq vaqt davomida farzand ko'rmagan ba'zi ayollar muqaddas joylarni ziyorat qilishadi. Bu tashrif buyurilgan muqaddas joyning nomi bolaga ism sifatida berilishi mumkin. Agar bolaning aniq ismi bo'lmasa, unga Nezir, Eren, Dede, Derviş, Adak kabi ismlar berilishi mumkin. Bu nom berish an'anasi qadimgi turk jamiyatlaridan beri davom etib kelgan.

Agar "Tabiat elementlari" deganda olov, suv, havo/shamol va yerni nazarda tutiladi va bu elementlar ham ism tanlashga motiv bo'lib xizmat qilishi mumkin. O'zbekistonda va Turkiyada ba'zi elementlar boshqalardan ko'ra bolalar uchun ism sifatida mashhurroq deb aytishimiz mumkin: Olov: Ateş, shuningdek, Alev (olov), Yalim (olov), Duman (tutun), Esan (shamol) o'g'il bolalar va qizlar (Alev) uchun ismlar sifatida juda keng tarqalgan. Kıvılcım (uchqun) qizlar uchun ham mashhur, o'g'il bolalar uchun kamroq.

Turkcha shaxs ismlariga qaraganda qimmatbaho metallar nomlari, (Zumrud, Yoqut, Feruza), rang nomlari va hokazolarni ham uchratishimiz mumkin.

Suv: Su (qizlar uchun, ba'zida o'g'il bolalar uchun ham), shuningdek, Irmak va Nehir (daryo), Deniz va Derya (dengiz), Yağmur (yomg'ir), Gursel (suv toshqini), Cansu (hayot suvi), Bora, Baran, Boran (yomg'ir, bo'ron), hatto Chag'layan (sharshara) kabi ismlar turklar orasida keng tarqalgan ismlar hisoblanadi va Suv bilan bog'liq boshqa o'nlab nomlar mavjud. O'zbeklarda ham suv bilan bog'liq Sharshara, Shalola, Daryo kabi ismlar tarqalgan.

Havo: Havo hech qachon o'g'il yoki qiz bolalar uchun ism sifatida ishlatilmaydi (Momo Havo bor, lekin bu Momo Havoni anglatadi - Xudo tomonidan yaratilgan birinchi ayol). Shamol (shamol), aksincha, kamdan-kam bo'lsa-da, nom sifatida ishlatilgan. Shuningdek, Poyraz (shimoli-sharqiy - sovuq shamol) - o'g'il bolalar uchun, Meltem (yumshoq shabada, zefir - qizlar uchun) juda mashhur. Hatto Tayfun (To'fon) ham o'g'il bolalar uchun mashhur ism hisoblanadi. Havo yoki shamol bilan bog'liq bir nechta boshqa nomlar mavjud.

Yer: O'g'il yoki qiz bolalar uchun ism juda kam uchraydi, lekin Turkiyada Toprak ismini keng tarqalgan ism sifatida qayd etish mumkin. Yer bizning sayyoramiz ma'nosini ham bildiradi va turkchada Yerning o'sha ma'nosiga mos keladigan ot, ya'ni Dunyo ismini uchratamiz

Kimyoviy elementlarga kelsak, bir nechta metallarning nomlari, ayniqsa Temir/Demir va Oltin/Altun (Oltin), Kumush/Gümüş va Bakir/Bakır (Mis) Feruza/Firuza (qimmatbaho tosh), Olmos/Almas (qimmatbaho tosh) kabi ma'danlarning nomlaridan ism sifatida foydalanish har ikkala xalq orasida birdek keng tarqalgan deyishimiz mumkin.

Dunyo tomonlari, yil fasllari va oylarning nomlarini ism sifatida qo'llash ko'proq turklarga xosdir. Quyidagi jadvalda ulardan namunalar ko'rishimiz mumkin

O'zbek	Turk	Ma'nosi
Yozberdi	Yaz	yoz faslida tug'ilgan qiz bolalarga beriladi
Yozpo'lat	Yozpolat	yoz faslida tug'ilgan o'gil bolalarga beriladi
Bahor	Bahor	bahor faslida tug'ilgan qizlarga beriladi
	Pınar	bahor degan ma'noni anglatadi
	Dogu	Sharq
	Bati	g'arb
	Kuzey	Shimol

	Gunay	Janub
	Evren	Koinot
Sayyora	Sayyara	Sayyora
	Uzay	Koinot
	Nisan	Aprel
	Haziran	Iyun
	Eylül	Sentyabr
	Ekim	Oytyabr
	Kasım	Noyabr

Ba’zi holatlarda farzandlar uchun ism nafosat, nur, ravshanlik, ma’nolarini ifoda etuvchi osmon jismlari nomidan ta’sirlanib ism tanlanadi: Kunbeka, Gunduz, Hilal, Bilal, Oycho’lpon, Oyyulduz, Aytekin (oy shahzodasi ma’nosida o’g’il bolalarga beriladi) Aytek (oy kabi), Aylin (oydan kelgan), Dolunay (to’lin oy), Aykız (oy kabi go’zal qiz)

Haqiqatan ham, aniqlikda professional darajaga yaqinlashadi degan fikr hayolliy bo’lib qolmoqda. Biroq, yuqori sifatli mashina tarjimasini olish samaradorlikni oshiradigan tadbirlar amalga oshirilmoqda

Foydalanilgan adabiyotlar:

1. Бегметов Е. Ўзбек тили антропонимикаси. 2013. 178 бет
2. İlknur Doğru “Türk kültüründe ad koyma olgusu ve ilgili inanışlar, 2020
3. Ермолович Д. И. Имена собственные на стыке языков и культур. М.: Р. Валент, 2001. С. 9
4. Литература и язык. Современная иллюстрированная энциклопедия / под ред. проф. А.П.Горкина. М.: Росмэн, 2006. 600 с.
5. Hurskainen A. Handling proper names in Machine Translation // Helsinki: Technical Reports in Language Technology Report. 2013, №12, P. 1–5.
6. Matthews D. Machine Transliteration of Proper Names. School of Informatics University of Edinburgh. 2007. P. 2.

BOLALARNING MATN PERSEPSIYASIDA SO‘ZLARNING ESLAB QOLINISHI**Abduraxmonova Sohiba Aktam qizi**

Mirzo Ulug‘bek nomidagi

O‘zbekiston Milliy universiteti magistranti

Annotatsiya: Ushbu maqolada psixolingvistikada bolalar nutqining o‘rganilishi, ularning nutqida matnni tushunish jarayonining bosqichlari haqida gap boradi. Shuningdek, bu jarayon tajriba amaliyoti orqali dalillab berilgan. Bugungi kunda psixolingvistikada bolalar nutqini o‘rganish dolzarbligi ochib berilgan.

Kalit so‘zlar: Matn darajasi, persepsiya, bola nutqi, dekodlash.

XIX asrning 60-70-yillarida psixolingvistikaning mustaqil fan sifatida shakllanishi bolalar nutqini o‘rganishga yanada keng imkoniyat yaratib berdi. Shu vaqtdan boshlab G‘arb tilshunosligida bola nutqini kuzatish orqali turli tajribalar asosida tadqiqotlar qilindi. Jahon tilshunosligida J.Berko, R.Braun, E.Klark, E.Beytes, M. Perroni, U.Stefani kabi olimlar tomonidan bolalar nutqidagi til o‘zlashtirish jarayonlari turli tillarda o‘ziga xos tajribalar orqali o‘rganilgan.

Rus tilshunosligida A.Leontyev⁸, L.Vigotskiy⁹, A.Shaxnarovich¹⁰, A.Gvozdev¹¹ kabi tadqiqotchilar bolalar nutqida til o‘zlashtirish jarayonlari, ona tilini o‘zlashtirishda psixolingvistik muommolar kabi jarayonlarni o‘rgangan. Qozoq tilshunosligida esa A.Aldaberdikizi¹², R.Abdiyeva, J.Nogaybayeva, A.Shinibayevalar¹³ qozoq tilining ona tili sifatida o‘zlashtirilishi, shuningdek qozoq tilining ikkinchi til sifatida o‘zlashtirilishi jarayonlarini o‘rganib chiqqan. Shuningdek, bolalar nutqi sintaksisi, fe‘l shakllarining o‘zlashtirilishi jarayonlarini T.Ayapova, D.Aldabergenova tomonidan bir qancha tadqiqotlar o‘tkazilgan.

O‘zbek tilshunosligida O.Umarxo‘jayeva, N.Sayidiraximova, G.Iskandarova, M.Qurbonova, va O.Saidaxmedovalar tomonidan bolalar nutqini qator tadqiqotlarda kuzatilgan va o‘zbek tilining birinchi til sifatida fonetik, grammatik xususiyatlari lingvistik, sotsiolingvistik, pragmalolingvistik va psixolingvistik jihatdan tadqiq etilgan.

Matnning antropotsentrik mohiyati psixologiya va lingvistika sohalari kesishuvi natijasida yuzaga kelgan psixolingvistikada, ayniqsa, yaqqol namoyon bo‘ladi. Psixolingvistik tadqiqotlarda matnning matn tuzuvchi – matn – retsiyentdan iborat uchlik¹⁴ nuqtayi nazaridan tadqiq etilishi unda shaxs omilining rolini chuqur o‘rganishni taqozo etadi.

Inson tashqi muhitdagi nutq tovushini o‘z lisoniy tajribasiga ko‘ra tushunishga harakat qiladi. Inson notanish tilga oid nutqni eshitganda ham undagi tanishday ko‘ringan so‘zlarga, fonemalarga yoki ohangiga ko‘ra xulosa chiqarishga urunadi. Tovushlarning mexanik qabuli doimo uning mazmuni haqida xulosa chiqarish bilan parallel kechadi. Tushunish darajalarining

⁸ Леонтьев А. А. Язык, речь, речевая деятельность. – Москва, 1969.

⁹ Выгодский Л. С. Мешление и речь. – Москва, 1982.

¹⁰ Шахнарович А.М. Психоллингвистические проблемы овладения общением в онтогенезе. – Москва, 1979.

¹¹ Гвоздев А. Н. От первых слов до первого класса. – Саратов, 1981

¹² Aldaberdikizy A. Lexical interferences of kazakh – Russian bilingual children of preschool age: A sociolinguistic survey // Procedia – Social and Behavioral sciences (2013). – 1885-1888.

¹³ Abdeyva R.S., Nogaybaeva J.A., Shynbaeva A.S. Kazakh language acquisition and translation from a foreigners perspective : a case study with English speaking residents // <https://www.science-sd.com>

¹⁴ Ковшиков В.А., Пухов В.П. Психология. Теория речевой деятельности: Учебник для вузов. – М.: Астрель, АСТ, 2007. – С.318

ajratilishi turlicha.V.P.Belyanin tushunish umumiy holatda uch darajadan iborat ekanligini ta’kidlaydi:

1.Boshlang‘ich daraja eng umumiy daraja bo‘lib unda nutqiy bayonning asosiy predmeti,gap nima haqda ekanligi tushuniladi.

2.Keyingi, mazmun mundarijani tushunish darajasi tinglovchi so‘zlovchining nima haqida nimalar deganini anglaydi.

3.Uchinchi daraja eng yuqori daraja bo‘lib , u qanday gap, nima uchun vaqanday lisoniy vositalar bilan aytilgani tushunish bilan xarakterlanadi¹⁵.

Matnni o‘qib tushunish murakkab psixolingvistik jarayon bo‘lib, u tilshunos olimlar tomonidan bir qancha bosqichlarga ajratiladi. Jahon tilshunosligida T.A. Apollonskaya, Y.V. Gleybman, I.Z. Manoli kabi olimlar qabul qilishning 3 ta asosiy darajasini belgilaydi:

- verbal-sintagmatik daraja;
- designativ daraja;
- dinatativ daraja.

O‘zbek tilshunosligida esa olima I.Azimova matnni tushunishning beshta darajasini keltiradi:

- a)assotsiativ daraja;
- b) leksik-morfologik daraja;
- c) kontekstual daraja;
- d) struktural daraja;
- g) matn darajasi¹⁶.

Hammada ham matnni eshitib yoki o‘qib tushunish bosqichlar bir xil amalga oshavermaydi. Shuning uchun ham muayyan matnni bir xil o‘qib yoki eshitgan o‘quvchilar tushunishining har xil bo‘lishini guvohi bo‘lamiz. Bu uning fiziologik va psixologik jihatlariga bog‘liq bo‘ladi, ya’ni qabul qilish qobiliyati yoki ruhiy-psixik holatiga, yoki jismoniy holatlariga bog‘liq bo‘lishi mumkin.

Tajriba materiali: Tajriba materiallari Davlat ta’lim standarti talablariga ko‘ra, o‘quvchilarning yoshiga, ya’ni 10-12 yosh o‘quvchilarining imkoniyatlari, qiziqishlari, shu paytga qadar egallagan bilim, ko‘nikma va malakalariga mos holda ularning psixologik, ijtimoiy holatlarini inobatga olgan holda saralab olindi. Tajriba materiali asosini, maktabdarsligi uchun berilgan matn tashkil qiladi.

Tajriba matni:”Tabiat va biz” matni.5-sinf ona tili darsligi 28-bet.Matn 118 ta so‘z, sarlavha va 9 ta gapdan iborat.

Tabiat va biz

Bugungi kunda tabiat insonning yordami va g‘amxo‘rligiga muhtoj ekanini barchamiz yaxshi bilamiz.Buning boisi – tabiatga g‘ayriinsoniy munosabat.Uzoq yillar davomida ishlab chiqarish munosabatlarining noto‘g‘ri tashkil etilgani tufayli ko‘plab ekologik muommolar vujudga keldi.

Ekologik muommolar tabiat qonuniyatlariga ko‘r-ko‘rona aralashish oqibatida kelib chiqadi.Masalan, yirik zavod, fabrikalardan turli zararli gazlarning chiqishi natijasida atmosferdagi ozon qatlami yemiriladi, hosil bo‘lgan tuyruk hozirgi paytda butun insoniyatni

¹⁵ Белянин В.П.Психоллингвистика. – М.,2001.-С.103

¹⁶ Azimova I.O‘zbek tilidagi gazeta matnlari mazmuniy persipsiyasining psixolingvistik tadqiqi.monografiya,T.:2019

tashvishga solmoqda. Buning natijasida quyoshdan keladigan turli zararli nurlar yemirilgan ozon qatlaminig tuynugi orqali tozalanmay, tog‘ridan-to‘g‘ri yerga tushmoqda.

Bilasizki, ona yurtimiz-dunyoda tengsiz go‘zal makonlardan biri. Purviqor tog‘lari, daryo va ko‘llari, musaffo osmoni, yam-yashil dalalari, bir so‘z bilan aytganda, butun tabiati har qanday insonni maftun etadi. Uni asrab-avaylash, kelajak avlodga bus-butun yetkazish siz va bizning qo‘limizda.

(“Tabiatni asrash o‘z qo‘limizda “ risolasidan)

Tajriba ishtirokchilari: 2022-2023- o‘quv yilida Toshkent shahri Olmazor tumani 165-maktabning 5-“g” sinfnig 10 nafar o‘quvchisi.

Tajriba usuli: Bu tajriba 3 qismdan iborat bo‘lib dastlab matn tadqiqot ishtirokchilariga o‘qib eshittirildi. Birinchi qismda o‘quvchilar matndan eslab qolgan qismidan so‘zlar ro‘yxatini tuzish, ikkinchi qismda eshitib tushunilgan matnga bayon yozib berish, uchinchi qismda esa matnni og‘zaki bayon qilib berish vazifasi topshirildi. Bolalarda eshitib tushunishni aniqlashdan maqsad dekodlash jarayonini aniqlash. Bilamizki, eshitish orqali tushunish sekinroq amalga oshadi. Ya‘ni ko‘z matndagi so‘zlarni ketma-ket qabul qilib borar ekan, ayni vaqtda ma‘lum so‘zni qabul qilish bilan birga atrofdagi so‘zlarni ham hisobga oladi¹⁷.

Birinchi qism bo‘yicha matn uch marta o‘qib eshittirilgach ishtirokchilar har bir o‘qishdagi eslab qolgan so‘zlarini qayd etdilar. Matnni har o‘qiganda eslab qolish samaradorligining oshishi persepsiya jarayonini ko‘rsatib beradi.

Matn mazmuniy persepsiyasida so‘zlarning eslab qolinishi

Matnni eshitib qabul qilishda so‘zlarning eslab qolinishi: Tajribada ishtirokchilar matnni eshitish orqali eslab qolgan so‘zlari ro‘yxatini tuzishdi. Sinaluvchilar matnni uch marta eshitgach eng kam 24ta eng ko‘p 45 ta so‘zni eslab qolganlar. Shuningdek, tajriba qatnashchilari ichida matnda mavjud bo‘lmagan so‘zlarni ham yozish kuzatilgan.

Masalan, 1-raqamli ishtirokchining eslab qolgan so‘zlarini quyida jadvalda ko‘rishimiz mumkin.

1-eshitishdan so‘ng	2-eshitishdan so‘ng	3-eshitishdan so‘ng
bugun	Kunda	insonning
barchamiz	Muhtoj	davomida
notog‘ri	uzoq	aralashish
tabiat	ko‘plab	hosil
daryo	ko‘rona	bilasizki
yurtimiz	Gazlar	musaffo
ko‘llar	yerga	ozon qatlami
tog‘	nur	yemirilyapti
osmon	yam-yashil	yirik
quyosh	Tuynuk	zavod
	ishlab chiqarish	fabrika
	ona vatan	chiqarayotgan

¹⁷ Azimova I. O‘zbek tilidagi gazeta matnlari mazmuniy persipsiyasining psixolingvistik tadqiqi. monografiya, T.: 2019

	qora tutunlar	hosil bo‘ladi
		zararli gazlar
		havoni zaharlayapti
		ob-havomizga ta’sir qilyapti
		tog‘lardan suv kelib

Bu tajribada 1-eshtitishda ishtirokchilar ma’lum bir so‘zlarni yozganlar keyingi eshitishda esa so‘zlar soni oshgan. Ushbu jadvalda ham qatnashuvchilardan birining so‘zlari ro‘yxati aks etgan. Bunda so‘zlar alohida shaklda esda saqlab qolingani, ikkinchi qismda esa so‘z birikmalari shaklida, shu bilan birga ketma-ketlik asosida esda qolgan. Uchinchi eshitishda ketma-ketlik tartibi oshgan va gaplar ham eslab qolingani. Ayrim sinaluvchilar matnda mavjud bo‘lmagan so‘zlarni ham yozishgan. Masalan 5-raqamli sinaluvchi birinchi eshitishda go‘zal, zararli, tuyruk so‘zlari o‘rniga chiroyli, xavfli, teshik kabi sinonimlar yozgan. Ikkinchi va uchinchi eshitishda ona yurtimiz, tashvishga solmoq so‘zlari o‘rniga ona vatanimiz, havotirga solmoq kabi so‘z birikmalardan foydalangan. Bu holatni ishtirokchida stimul matn ta’sirida yuzaga kelgan assotsiatsiya deyish mumkin.

Sinaluvchilar tomonidan yaxlit eslab qolingani hamda ularning bayonlarida qo‘llangan so‘zlar zanjirining statistikasi:

Yaxlit eslab qolingani so‘zlar zanjiri va gaplar	Sinaluvchilarning soni
bugungi kunda tabiat	4
ekologik muommolar vujudga keldi	3
zararli gazlar	6
ozon qatlami yemiriladi	5
insoniyatni tashvishga solmoqda	3
to‘g‘ridan-to‘g‘ri yerga tushmoqda	2
go‘zal makonlardan biri	4
insonni maftun etadi	4
asrab –avaylashimiz kerak	7
uzoq yillar davomida	2
aralashish oqibatida	1
tashvishga solmoqda	2
yirik zavod va fabrikalar	6

Matnning qanday tushunilganini tahlil etishda biz stimul matn bloklarining qanchalik ifodalanganida ko‘rishimiz mumkin. Tajribada ishtirok etgan 10 nafar ishtirokchining 4 tasi matn mazmunini yaxshi tushungan. Matn mazmunini bayonda to‘liq ko‘rishimiz mumkin. Quyida 2-raqamli sinaluvchining bayoni aks etgan.

Tabiat va biz (A.I)

Bugungi kunda tabiat insoniyatning yordamiga muhtoj. Ancha yillardan beri qurilgan fabrikalar tabiatga zarar yetkazayapti. Masalan: zavodlarning zararli tutuni koinotdagi yer azon qatlamini yedirayapti. Undan hosil bo‘lgan teshik hozirda odamlarni vahimaga solyapti. Ba’zi bir quyoshning zararli nurlari to‘g‘ridan-to‘g‘ri azon qatlamining teshigidan yerga

tushyapti. Yerimiz beqiyos tog'lar, musaffo suvlar, moviy osmon va o'simliklarni asrab kelajak avlodga yetkazib berish sizubizning vazifamiz.

Tajriba shuni ko'rsatdiki, sinaluvchilar yozgan matnlar ayni matndan shakliiy belgilariga ko'ra farqlanadi. Ayrim so'zlar o'rniga sinonimlaridan foydalanish, aniqllovchilarning tushib qolishi, bitta gapning ikkita gapga aylanishi yoki gaplarning bitta gapga aylanishi hodisalarini kuzatdik. Shuningdek, qatnashuvchilar matnda mavjud bo'lmagan so'zlarni ham yozishgan.

Xulosa sifatida shuni aytish mumkinki, bolalarda matn persepsiya jarayoni turlicha kechadi. Bu ularning aqliy va psixologik holatlari bilan bog'liq. Qaysidir bola matnni bir eshitganda qabul qilsa, boshqa bola uchinchi eshitishda ham matn haqida muayyan xulosaga kela olmasligi mumkin. Bir so'z bilan matn persepsiyasini individual jarayon deb aytishimiz mumkin.

Foydalanilgan adabiyotlar:

1. Леонтьев А. А. Язык, речь, речевая деятельность. – Москва, 1969.
2. Выгодский Л. С. Мешление и речь. – Москва, 1982.
3. Шахнарович А. М. Психоллингвистические проблемы овладения общением в онтогенезе. – Москва, 1979.
4. Гвоздев А. Н. От первых слов до первого класса. – Саратов, 1981
5. Aldaberdikzy A. Lexical interferences of kazakh – Russian bilingual children of preschool age: A socialinguistic survey \ Procedia – Social and Behavioral sciences (2013). – 1885-1888.
6. Ковшиков В. А., Пухов В. П. Психология. Теория речевой деятельности: Учебник для вузов. – М.: Астрель, АСТ, 2007. – С. 318
7. Беянин В. П. Психоллингвистика. – М., 2001. – С. 103
8. Azimova I. O'zbek tilidagi gazeta matnlari mazmuniy persipsiyasining psixolingvistik tadqiqi. monografiya, T.: 2019

4-SHO'BA

KOMPYUTER LEKSIKOGRAFIYASI: ONTOLOGIYA, TEZAURUS

MA'LUMOTLAR BAZASINI SHAKLLANTIRISHDA SO'ZLARNI TASNIFLASH MUAMMOLARI

Qo'ziyev Umidjon

PhD, dotsent. Namangan davlat universiteti

Annotatsiya: Ushbu maqolada o'zbek tili milliy korpusini yaratishda, so'zlarni to'plamlarga biriktirishda yuzaga keladigan muammolar yoritilgan. Yechim sifatida to'plamlar, so'z turkumlari tarkibi va tavsifini o'zbek tilining tabiatidan kelib chiqib qaytadan ko'rib chiqish tavsiya qilinadi.

Kalit so'zlar: korpus, to'plam, so'z turkumlari, ravish, o'zlashma so'zlar

Abstract: This article describes the problems in the composition of the national corpus of the Uzbek language, and in connecting words to parts of speech. As a solution, it is recommended to reconsider the composition and description of parts of speech, word groups based on the nature of the Uzbek language.

Keywords: corpus, collection, word groups, idioms, borrowed words

Аннотация: В данной статье описаны проблемы, возникающие при создании национального корпуса узбекского языка, и при соединении слов в части речи. В качестве решения рекомендуется пересмотреть состав и описание сборников, словосочетаний исходя из особенностей узбекского языка.

Ключевые слова: корпус, коллекция, словосочетания, идиомы, заимствованные слова.

O'zbek tili milliy korpusini yaratishda, so'zlarning grammatik bog'lanishlari, assotsiatsiyalari to'g'ri kiritish, bunda tabiiy tildagi xususiyatlari to'liq hisobga olishimiz kerak bo'ladi.

Hozirda o'zbek tilidagi matnlarni imloviy tahrir qilish borasida yutuqlarga erishilganiga qaramay, morfologik va sintaktik tahlil masalalari oqsamoqda.

Bunga sabab sifatida o'zbek tilining murakkab tabiatli ekanligi, ularni kiritish uchun axborot ombori to'liq emasligi, o'zbek tili leksik tarkibi korpusga to'liq yuklanib ulgurmaganligi bilan ham izohlanishi mumkin. Biroq bu o'rinda tabiiy tilga oid grammatik qoidalarning o'zi murakkab emasmi, bu qoidalar asosida ma'lumotlar bazasini shakllantirish ko'plab muammolarni keltirib chiqarmayaptimi? degan savol tug'iladi.

Morfologik jihatdan o'zbek tili – agglyutinativ til. Agglyutinativ tilning asosiy jihatlari sifatida quyidagilar ko'rsatiladi:

1. So'zning grammatik shakllari asosan qo'shimchalar – affikslar bilan hosil qilinadi.
2. Affikslar odatda suffiks xarakterida bo'lib, o'zakdan so'ng qo'shiladi.

3. Qo‘shimcha qo‘shilganda o‘zak va qo‘shimchaning o‘rni saqlanadi, ya’ni fleksiya yuz bermaydi.

Biroq biz agglyutinativ til deb baholayotgan o‘zbek tilida shu qoidalarining ham to‘liq tasdig‘ini ko‘rmaymiz. Birinchidan, affikslardan tashqari o‘zbek tilida yordamchi so‘zlar (ko‘makchilar kabi), affiksoidlar guruhi ham mavjud. Ikkinchidan, ko‘pgina so‘zlar yasashida fors-tojik tilidan o‘zlashgan prefikslar ham kuzatiladi: bama’ni, betaraf kabi. Uchinchidan, qo‘shimcha qo‘shilganda tovush ortish, tushish, o‘zgarishi kabi hodisalar ko‘p uchraydi.

Muammo nimada?

Muammo bizning tilimizda emas, bizning o‘z tilimiz haqidagi noto‘g‘ri tushunchamizda deb bilamiz. Biz o‘z tilimizni boshqa millat vakillari ko‘zi bilan tuzilgan grammatika orqali o‘qiymiz. Natija esa grammatikada qoidalarining istisnolarga to‘lib ketganligi, kompyuter lingvistikasida esa tilimizni ma’lum barqaror qoidalarga bo‘ysundirib bo‘lmayotganida yaqqol ko‘rinib turibdi.

Professor Abdumajid Po‘latov o‘zbek tilining formal grammatikasini tuzishni taklif qilganida, balki, shu qiyinchiliklarni ham nazarda tutgandir.

Yuqoridagi masalalar bo‘yicha o‘zimizning ayrim takliflarimizni bermoqchimiz.

O‘zbek tili fonemalar tarkibini qadimgi turkiy tildan o‘stirib qayta ko‘rib chiqish lozim. Shunda haligacha o‘gay bo‘lib turgan –ng birikmasi yoniga –nch, –ny kabi fonemalar ham qo‘shilib, ular uchun alohida grafemalar olinardi. Natijada

u+ga – unga hosilasida tovush orttirilgan bo‘lmaydi, balki *u-nga*, shu-nga tarzida o‘rtacha yumshoqlikdagi *ng* ni ifodalovchi harf muomalaga kiritilardi.

Natijada

U-ncha, shu-ncha, ikki-nchi, sev-inch kabi hosilalarda mushtaraklik hosil bo‘lar va sun‘iy tovush orttirilishi yukidan xalos bo‘lar edik.

Biroq *-da, -dan* kabilarni qo‘shishda yuzaga kelayotgan *n* ortishini qayta ko‘rib chiqish lozim bo‘ladi. Ikkinchidan, *me-nga, se-nga* kabi holatlar uchun o‘zak sinxron planda ma’no bildirmaydimi yoki menga, senga xuddi ingliz tilidagi *me*, rus tilidagi *mne, menya* kabi alohida so‘z maqomiga ega bo‘ladimi?

Keyingi yirik masala – so‘z turkumlarining to‘liq tasniflanmagani bo‘lib, natijada grammatikamizda ravish, modal so‘z, yuklama singari ta’rifi o‘ziga mantiqan mos kelmaydigan sun‘iy so‘z turkumlari yuzaga keltirilgan.

Ravish so‘z turkumini oladigan bo‘lsak, qisqa qilib aytganda, “harakatning belgisini bildirib, o‘zgarmaslik xarakteriga ega bo‘lgan” so‘z turkumi sifatida ta’riflanadi.

Muammo: ravishdagi “harakatning belgisi” ta’rifi so‘zning bosh UGM sini emas, 3-darajali – sintaktik xususiyatini ifodalashi ma’lum bo‘ladi. YA’ni boshqa so‘z turkumlari ham fe’lga bog‘lanib, uning belgisi – payt, sabab, maqsad kabilarni bemalol ifodalay oladi.

1-jadval

	Ravish*	Ot	Sifat	Son
Payt	<i>Kecha, bugun, oldin, keyin...</i>	<i>Juma kuni, sentyabrda</i>	–	<i>Soat 5 da, o‘n beshu nol nolda</i>
O‘rin	<i>Oldinda, ilgari, olg‘a</i>	<i>Uyda...</i>	–	–

Sabab	Noiloj, noilojlikdan	Soddaligidan	–	–
Maqsad	Atay, ataylab	Kitobga, o'qishga	–	–
Holat	Tez, sekin, asta	Tezlik bilan, ohistalik bilan	Yaxshi, qulay	–
Daraja-miqdor	Ko'p, oz, juda	–	–	Yuzta

Qoida bo'yicha, ravish yuqoridagi ma'nolarni, boshqa turkumlardan farqli ravishda qo'shimchasiz, shakllarsiz bajara olishi kerak. Biroq kechadan, bugunga, ilgaridan kabilarda kelishik shakllari bilan turlanmoqda. Demak, bu jihatdan ravish boshqa turkumlardan afzal bo'lmaydi.

Ikkinchi jihat ravishlarning "o'zgarmasligi" haqidagi mif bo'lib, bu mifga asosan ravishlar egalik, kelishik shakllarini ololmaydi. Biroq yuqoridagi jadvaldagi ravish qatorida berilgan so'zlarning 80 foizi turlanish xususiyatiga ega.

Bu holatga BuxDU dotsenti T.Asadov ham e'tibor berib, ravish so'z turkumi mavjudligini, biroq nisbatan keyingi davrlarda yuzaga kelganini qayd etadi.

Unda qayerda xato qilingan?

Fikrimizcha, oldingi so'z turkumlarining tahlili, tasnifi to'liq amalga oshirilmagan. Natijada bu turkumlardan ortib qolgan so'zlar qayta tekshirib ko'rilmadan, ravish turkumiga surib tashlangan.

Ot so'z turkumidan misol ko'ramiz:

2-jadval

Semantik turi	Atoqli ot		Turdosh ot	
	aniq	mavhum	Aniq	Mavhum
Predmet	"Matiz"	Ekskalibur	mashina	Eleksir
SHaxs	Bobur	Gorgona	Bola	Farishta
O'rin-joy	Olmalik	Jannat	shahar	???
Payt	Navro'z	Qiyomat	Kuz	???
Faoliyat-jarayon	"Obod mahalla"	???	Terim	muhabbat

E'tibor bersak, o'rin-joy bildiruvchi mavhum otlarga tepa, past, yuqori, quyi; payt bildiruvchi mavhum turdosh otlarga oldin, keyin, so'ng kabi ravish*lar bemalol misol bo'la oladi.

Natijada bu so'zlarning turlanishi haqidagi istisno qoidalar bo'lmaydi.

Bundan tashqari oz, ko'p kabi so'zlarni miqdor bildiruvchi sifatlarga, juda, g'irt kabilarni yuklamalarga olish mumkin. Shunda ravishlarning daraja bildirish uchun o'zidan oldin o'zini keltirishdek qoidasidan voz kechamiz: juda tez.

So'zlarning sintaktik munosabatlarini kiritishda shu kabi notabiiy istisnolarning kamayishi ma'lumotlar bazasining mantiqan to'g'ri shakllantirilishi, dasturiy ta'minot hajmining soddalashishiga sabab bo'ladi.

Foydalanilgan adabiyotlar:

1. Асадов Т.Х. Сўз туркумлари тизимида равиш. Монография. – Тошкент: Мухаррир, 2010. 132 б.

2. Искандарова Ш. Систем тилшунослик масалалари. Монография. – Тошкент: Фан, 2007. 54 б.
3. Po‘latov A., Muhamedova S. Kompyuter lingvistikasi. – Toshkent, 2009.
4. Ҳожиёв А. Ўзбек тилида равиш ясаиши тизими масаласига доир //Ўзбек тили ва адабиёти, 2001, №2, 42-46- б.
5. Ҳожиёв А. Ўзбек тили морфологияси, морфемикаси ва сўз ясаишининг назарий масалалари. – Тошкент: Фан, 2010. 256 б.
6. Ҳожиёв А. Ўзбек тили сўз ясаиш тизими. – Тошкент: Ўқитувчи, 2007. 170 б.
7. Kuziyev U. About bases of classification of the words in composing new complanatory dictionaries //Scientific journal of the Fergana State University. – 2018. – Т. 1. – №. 1. – С. 104-106.
8. Qo‘ziyev Umidjon Yandashaliyevich. (2023). O‘Z QATLAMDAgi O‘ZLASHMA QOLIPLAR. "Uchinchi Renessansda Ilmiy-Amaliy Tadqiqotlarning Dolzarb muammolari" Mavzusidagi Onlayn Konferensiyasi, 1(8), 112–115. Retrieved from <http://re-search.uz/index.php/conf/article/view/166>

TIL BIRLIKLARNING SEMANTIK XUSUSIYATLARI

Israilova Saodat Turapovna,

Senior teacher of National University
of Uzbekistan candidate of philological sciencesE-mail: israilovas404@gmail.com

Annotatsiya: Sintagmatik munosabatlar shunday munosabatlarki, ularga ko'ra til aloqa vositasi sifatida xizmat qiladi, ya'ni, tilning kommunikativ vazifasi sintagmatik munosabatda namoyon bo'ladi va shu munosabatga tayanadi. Paradigmatik va sintagmatik munosabatlar o'zaro dialektik bog'liqlikda, bir butun, yaxlit holda sistemalar umumiyliigi, jami sifatida tilning aloqa quroli bo'lib xizmat qilishini, ijtimoiy-amaliy faoliyatini ta'minlaydi.

Kalit so'zlar: sintagmatika, paradigmatic munosabatlar, nutqiy birliklar tahlili, ko'k yadroli birliklar, semantik ma'no.

Abstract: Syntagmatic relations are such relations according to which the language serves as a means of communication, that is, the communicative function of the language is manifested in the syntagmatic relation and relies on this relation. Paradigmatic and syntagmatic relations are mutually dialectical, the totality of systems as a whole, as a whole, ensures that the language serves as a tool of communication and social-practical activity.

Keywords: syntagmatics, paradigmatic relations, analysis of speech units, blue core units, semantic meaning.

Аннотация: Синтагматические отношения - это такие отношения, согласно которым язык служит средством общения, то есть коммуникативная функция языка проявляется в синтагматическом отношении и опирается на это отношение. Парадигматические и синтагматические отношения взаимно диалектичны, совокупность систем в целом, в целом обеспечивает то, что язык служит орудием общения и социально-практической деятельности.

Ключевые слова: синтагматика, парадигматические отношения, анализ речевых единиц, синие стержневые единицы, семантическое значение.

Jamiyat rivojlanishining qizg'in pallasiga kirgani sari tilimizda o'ziga xos taraqqiyot yuz berishi tabiiydir. Binobarin, tildagi so'zlarning shakl va ma'no jihatdan taraqqiy etishi ulardagi ushbu munosabatlarni kengroq tadqiq etishni taqozo etadi. So'zlarning o'zaro shakl va ma'no munosabatlari turlicha. Ayrinim so'zlar shaklan har xil, lekin ma'no jihatidan bir-biriga yaqin bo'lsa, boshqalari shaklan bir xil, ammo ma'nolari turli xil, ba'zilar esa bir-biriga qarama-qarshi so'zlar ham bo'lishi mumkin.

Sintagmatikada lisoniy birliklar orasidagi munosabatlar aniqligi va izchilligi bilan ajralib turadi. Qiyoslang: *Ekilgan ko'chatlar viqor bilan ko'kka bo'y cho'za boshladi*. Ko'rinib turibdiki, nutq faoliyatida muayyan "yotiq chiziq"qa joylashib, muayyan tartibda, mantiqiy ketma-ketlikda kelgan birliklar orasidagi munosabatlar sintagmatik munosabatni hosil qiladi. Sintagmatik munosabatlar doimo yaqqol ta'sir qilishi □ insonda lisoniy sezgi uyg'otishi bilan ham o'ziga xosdir.

Sintagmatik munosabatlar nutq faoliyatidagi birliklar aloqasi, munosabati sifatida, paradigmatic munosabat birliklaridan farqli, bir guruh birliklari orasida ham, shuningdek, turli guruh birliklari orasida ham mavjud bo'lishi mumkin. Qiyoslang: *Ko'klam kirib keldi va hammayoq yashilga burkandi*. Ushbu birliklardan har biri o'zi tegishli bo'lgan guruhning-paradigmatikaning vakili hisoblanadi. Demak, sintagmatikadagi munosabatlar muayyan birliklar-so'zlar munosabati sifatida emas, balki guruhlar-sinflar, paradigmalar sistemasi munosabati sifatida namoyon bo'ladi.

Semantik paradigma ma'no shakllanish jarayoni haqida ma'lum ilmiy tasavvur beradi. "Ko'k" leksemasi ishtirokidagi iboralar tahlili orqali leksemaning semantik imkoniyatlarini unga yondosh, u bilan bog'langan boshqa birliklar semantikasi orqali ochib berishga harakat qilamiz.

"Ko'k" leksemasi bir qator kasallik nomlariga ham asos bo'lgan:

Ko'ksov – o'pkasi ishdan chiqqan, ko'p va qattiq yo'taladigan kasallik belgisi. Ko'p hollarda hayvonlarga nisbatan qo'llanadi: *Chang orasida bitta-yarimta ko'ksov qo'ylarning yo'talgani eshitilar edi*. G'.G'ulom, Shum bola (O'TIL,II,451).

Ko'ksuv – ko'zdan yosh (suv) oqib turishi bilan kechadigan ko'z kasalligi. (O'TIL,II,451).

Ko'kyo'tal – tib. O'qtin-o'qtin tutib, uzluksiz yo'taltiradigan, odatda bolalarda bo'ladigan yuqumli kasallik. *Yo'talning kechasi, shuningdek, dam-badam tutishi ko'kyo'talning alohida xususiyatlaridandir*. N.Ismoilov, Kasallarni parvarish qilish (O'TIL,II,451).

Ko'k yadroli iboralar ham tilimizda ko'plab uchraydi. **Boshi ko'kka yetmoq** iborasi insonning ruhiy holatini ifoda etadi, xursandlikning yuqori darajasini ifodalaydi. **Bosh metonimiya asosida inson ruhiyatini aks ettiradi, ko'k esa ruhiyatning quvonch ta'siridagi maqomini ifoda etadi.**

Boshi ko'kka yetmoq iborasi shaxs o'zida yuzaga kelgan ruhiy holatni aks ettirsa, "tavsifini juda oshirmoq, o'ta maqtamoq" **ko'kka ko'tarmoq** barqaror birikmasida "boshqa shaxs tomonidan amalga oshish" semasi reallashgan. Bu sema **ko'tarmoq** harakat fe'lining holat fe'li sifatida qo'llanishida namoyon bo'ladi. **Ko'k** maqtash, ulug'lash faoliyatining yuqori darajasini ifodalashga xizmat qilgan.

Yeru ko'kka ishonmaslik iborasi "o'ta darajada ehtiyotlamoq" ma'nosini bildiradi. "Ko'k" leksemasi *yer so'ziga* nisbatan antonimik munosabatni ifodalab, birgalikda "olam", "dunyo" semalarini anglatgan. O'z navbatida "hech qaysi joy" konseptini namoyon qilgan. Qayd etilgan iboralarda ijobiy semalar ifodalangan.

Ko'k yadroli quyidagi iboralarda esa salbiy sema, baho aks etgan:

Kulini ko'kka sovurmoq iborasi katta hajmdagi, keng ko'lamdagi voqelikka bo'lgan salbiy munosabatni ifodalaydi. Ibora "yo'qotish" semasini darajalab ko'rsatishga xizmat qiladi. Kulini ko'kka sovurmoq iborasida quyidagi mantiqiy-semantik jarayon o'z ifodasini topgan: 1.Olovda yoqish. 2.Kulga aylantirish. 3.Kulni ko'kka sovurish – mayda zarrachalarga bo'lish, yo'qotish. **Ko'k** "osmon", "kenglik", "koinot" semalarini ifoda etgan. Iborda "tugallanganlik" semasi ham mavjud. Chunki **kul** olovning oxirigacha yonib bo'lish jarayoni natijasidir.

Tutuni ko'kka chiqmoq iborasi ham "olov" tushunchasi bilan bog'liq. "O'ta darajada xafa bo'lmoq, xunob bo'lmoq" ma'nolarini ifodalaydi. Iborda "tugallanmaganlik" semasi mavjud. Chunki tutun o'chgan olov belgisi bo'lib, jarayonning davomini ifoda etadi. Yonayotgan ashyoning kuyishidan dalolat beradi. Demak, ibora semantikasi ashyoning yona olmay kuyishi va

insonning o'zni qayta yondira olmay, asabiylashishiga asoslanadi. *Ko'k* bu ruhiy holatning yuqori darajasini ifodalashda markaziy o'rin tutadi.

Ko'k yadroli “*Ko'kning hidini olgan sigir boyloqda turmas*”, “*Ko'kka boqqan yiqilur, loy-botqoqqa tiqilur*”, “*Ko'k kirsin, qizil chiqsin*”, “*Ko'kdan arpa yog'sa ham, eshakning yemi – nimcha*” (nimcha – “nimxo'rd” ham derdilar, chorakning to'rtidan biri, botmonning esa 256 dan biri), “*Kekkayganga kekkaygin boshing ko'kka yetguncha, enkayganga enkaygin boshing yerga tekkuncha*”, “*Daraxt bir joyda ko'karadi*”, “*Dalaning ko'kiga ishonma, odamning – po'kiga*”, “*Ko'kini ko'rib, ko'ksingni kerma*”, “*Dalaning ko'kiga ishonguncha, otangning go'riga ishon*”, “*Duo bilan el ko'karar, yomg'ir bilan yer ko'karar*”, “*Duo olgan ko'karar, tuhmat olgan oqarar*”, “*Tinchlik bilan el ko'karar, yomg'ir bilan – yer*”, “*Ayol yerdan chiqqan emas, erkak bolasi, erkak ko'kdan tushgan emas, ayol bolasi*”, “*Aql ko'pga yetkazar, hunar – ko'kka*”, “*Odobli o'g'il – ko'kdagi yulduz, odobli qiz – yoqadagi qunduz*”, “*Oydinda oq kiygan, ko'lankada ko'k kiyar*” paremiyalari “belgi”, “osmon”, “yuksaklik” semalari asosida turli mental mazmunni ifodalashga yo'naltirilgan.

Xususan, qozoq tilida ham “*ko'k*” leksemasi bilan shakllangan quyidagicha milliy-madaniy birliklarni uchratamiz: “*Ko'k*” leksemasining “osmon” ma'nosidagi lisoniy birliklari: *kök jyzi* → osmon (*ko'k yuzi*); *kök kymbez* → falak; *kök qasqa aytu* → qurbon qilmoq; qurbonlik qilish (otni); *kök seңgir* → cho'qqi; *kök tireu* → juda baland ko'tarilmoq; osmonga egilib; *kökke boylau* → baland o'smoq (daraxtlar haqida); *kökke boylag'an qayyındar* → baland qayinlar; *kökke boylag'an mynaralar* → baland minoralar (*ko'kka cho'zilgan minoralar*); *kökke köteru* → maqto'v (*ko'kka ko'tarmoq*); *kökke köterilu* → baland uchmoq; ko'tarilish; *kökke qarau* → takabbur; o'zini boshqalardan ustun qo'ymoq; *kökke qylash yru* → yuksaklikka intilmoq; *kökke samg'au* → osmonga uchmoq; *kökte qalyıqtau* → osmonda uchmoq (qush haqida); *kökpen talasqan* → juda baland; *kökten tyskendey* → osmondan tushgandek, birdan; *olardıң arasy jer men köktey* → ular orasidagi farq osmon bilan yer orasidagi kabi.

Xulosa qiladigan bo'lsak, ko'k so'zidan shakllangan 25ta leksema hozirgi o'zbek tilida ishlatiladi. Ko'k ishtirok etgan 150 ta turli hajmdagi (2 ta so'z→37 ta so'z: eng kichigi ikkita so'z va eng kattasi 37 ta so'zdan iborat) matn namunalarini to'pladik. Tahlillar shuni ko'rsatdiki, mumtoz badiiy matnlarda ko'k leksemasining “osmon”, “balandlik”, “yuksaklik” semalari faol qo'llangan. Hozirgi davr matnlarida esa ko'k leksemasining “yashil” semasi dominantlik qiladi. Ko'pincha, ko'k so'ziga rang jihatidan qaralsa, qizil yo sariq so'zlari nutqimizdan kelib chiqqan holda qarama-qarshi qo'yiladi. Ko'k-qizil, ko'k-sariq kabi. Agar ko'k leksemasiga osmon, samo ma'nolari jihatidan qaralsa, ko'kka yer so'zi qarshi qo'yiladi.

Foydalanilgan adabiyotlar:

1. Axmanova O.S. Slovar lingvisticheskix terminov. – M., 1960. –S. 310.
2. DTS – Drevnetyurkskiy slovar. – L., 1969. – 677 c.
3. Etimologicheskii slovar tyurskix yazыkov. Obshetyurskie i mejtyurskie leksicheskie osnovы na bukvы “K”, “Q”. – M.: 1997. – 363 c.
4. Sevortyan E.V. Etimologicheskii slovar tyurskix yazыkov.VI. – Moskva: Nauka, 1974. – 767 c.

5. O‘zbek tilining izohli lug‘ati. – Toshkent: O‘zbekiston Milliy ensiklopediyasi Davlat ilmiy nashriyoti, 2006. I –680 b.; II– 672 b.; III – 688 b.; IV - Toshkent, 2008. -608 b.; V – 592 b.
6. O‘zbek tilining izohli lug‘ati. 5 jildli. 1-jild. – Toshkent: O‘zbekiston milliy ensiklopediyasi, 2006. – 680 b.

TURKIY TILLARDA CHASTOTALI LUG'AT TUZISH TARIXIDAN

Otemisov Aziz

PhD., kafedrası dotsenti,

Berdaq nomidagi Qoraqalpoq davlat universiteti

E-mail: otemisov.aziz@mail.ru

Аннотация: В статье рассматривается значение частотных словарей, играющих важную роль в компьютерной лингвистике, история создания частотных словарей в тюркских языках, частотных словарей, разработанных в казахском, узбекском, турецком, кыргызском и каракалпакском языках. Сегодня отмечается необходимость и актуальность частотных словарей при создании языковых корпусов.

Ключевые слова: компьютерная лингвистика, модель, частотный словарь, языковой корпус, машинный перевод.

Annotation: The article discusses the importance of frequency dictionaries, which play an important role in computational linguistics, the history of the creation of frequency dictionaries in the Turkic languages, frequency dictionaries developed in Kazakh, Uzbek, Turkish, Kyrgyz and Karakalpak languages. Today, the need and relevance of frequency dictionaries in the creation of language corpora is noted.

Keywords: computational linguistics, model, frequency dictionary, language corpus, machine translation.

Chastotali lug'atlar yordamida tadqiq qilinayotgan matnda ko'p va kam uchraydigan soz va so'z shakllarini aniqlash – tilshunoslikdagi ba'zi nazariy va amaliy masalalarni hal qilishda katta ahamiyat kasb etadi. Sababi, bu lug'atlarga asoslanib, tilning leksik-statistik modelarini tuzishda tilshunoslikdagi bevosita nazorat qilishda aniqlab bo'lmaydigan yashirin qonuniyatlar va xususiyatlarni topishda, shuningdek, orfografik lug'atlarni takomillashtirishda, mashina yordamida bir tildan ikkinchi tilga tarjima qilishda, tilni qisqa muddat ichida o'rganib chiqishga, qoraqalpoqcha-ruscha, ruscha-qoraqalpoqcha, qoraqalpoqcha-inglizcha, inglizcha-qoraqalpoqcha minimum lug'atlarni tuzishda v.h.kazo o'zining ijobiy tasrini ko'rsatadi [Насыров, Айымбетов, 1983; 3].

Chastotali lug'atlar matnda so'zlarning ishlatilish chastotasi (qay darajada ishlatilishi, tez-tez ishlatilishi)dan kelib chiqqan holda tayyorlanadigan so'zliklar hisoblanadi. Chastotali lug'atlar statistik leksikografiyaning rivoji natijasida yuzaga kelgan. Bunday lug'atlar, ayniqsa, til o'qitish jarayonida leksik minimumlarni aniqlashda muhim ahamiyatga ega. So'zlarning chastotasini aniqlash linvostatistik tahlillar yordamida amalga oshiriladi. Chastotali lug'atlar kompyuterning lingvistik informatsion bazasi hisoblanadi. Masalan, chastotali lug'atdagi birinchi 10 va 50 eng ko'p uchraydigan so'z formalarining miqdori umumiy so'z birliklariga nisbatan agglyutativ tillarda flektiv tillarga qaraganda kamroq foizni tashkil etadi. Agglyutativ tillarda dastlabki 1000 so'z formasi umumiy so'z birliklarining 50-60 foizini tashkil etsa, bu ko'rsatkich flektiv tillarda 62-87 foiz atrofida bo'ladi. kelgan. Bunday lug'atlar, ayniqsa, til o'qitish jarayonida leksik minimumlarni aniqlashda muhim ahamiyatga ega. So'zlarning chastotasini aniqlash linvostatistik tahlillar yordamida amalga oshiriladi. Chastotali lug'atlar kompyuterning lingvistik informatsion

bazasi hisoblanadi. Masalan, chastotali lug'atdagi birinchi 10 va 50 eng ko'p uchraydigan so'z formalarining miqdori umumiy so'z birliklariga nisbatan agglyutinativ tillarda flektiv tillarga qaraganda kamroq foizni tashkil etadi. Agglyutinativ tillarda dastlabki 1000 so'z formasi umumiy so'z birliklarining 50-60 foizini tashkil etsa, bu ko'rsatkich flektiv tillarda 62-87 foiz atrofida bo'ladi [Rahimov, 2011; 51].

Turkiy tillar bo'yicha dastlabki chastotali lug'atlar o'tgan asrning 70 - yillarding o'rtalaridan e'tiboran qo'lga olindi. O'sha davrdagi EVM ES elektron hisoblash mashinalari yordamida chastotali lug'atlar tuzila boshladi. Ilk marta turkiy tillar ichida qozoq tilining matnlari bo'yicha har xil turdagi lug'atlar tashkil etildi. Qozoq kompyuter lingvistikasining shakllanishida prof. Q.B.Bektaevning ilmiy faoliyati muhim ahamiyatga ega. Olim matematik va injener lingvistikasiga oid o'nlab asarlar yaratgan. Q.B.Bektaevning boshchiligidagi Qozog'iston Fanlar Akademiyasining tilshunoslik instituti yonidan "Lingvostatistika va avtomatlashtirish" guruhi tashkil etildi. Guruhining a'zolari M.Avezovning "Abay yo'li" romani (4 kitob) tilining chastotali lug'atini kompyuter yordamida 50 soat davomida tuzib chiqqanlar. Mazkur chastotali lug'atda 20000 dan ziyodroq leksema va 60000 so'z shakli ("glossema") mavjud bo'lib, u 466 000 marta qo'llanilgan. Guruh a'zolari bu lug'atni tuzish uchun atigi sakkiz oy vaqt sarflaganlar. Vaholanki, shu ish oddiy qo'l kuchi bilan bajarilganda, mazkur guruh a'zolari tinimsiz 10 yil, bir tilchi esa kamida 100 yil ishlagan bo'lar edi. Shu tariqa ushbu ulkan ish qozoq tilshunosligida kompyuter lingvistikasining rivojlanishiga asos bo'ldi [Rahimov, 2011: 60]. Bu guruh umumiy hajmi million so'z qo'llanilishi atrofida uch turdagi lug'atni tuzib chiqdi.

1. Alfavitli chastotali lug'at
2. Chappa alfavitli chastotali lug'at
3. Chastotali lug'at, rangli chastotali lug'at

Bundan boshqa qozoq gazetalarining matnlari bo'yicha umumiy hajmi 150 mingta so'z atrofida yuqoridagi qayd etilgan uchta lug'at A. Aqabaev tomonidan tuzib chiqildi. Bundan so'ng A. Muxammedovning muallifligida o'zbek gazeta matnlari bo'yicha umumiy hajmi 200 mingta so'z atrofidagi chastotali lug'atlar tashkil etildi. Bundan boshqa bunday turdagi lug'atlar turk tili (A. Bobonarov), qirg'iz matnlari bo'yicha (M. Jetekishov) ishlab chiqildi [Айымбетов, 2013, 16].

O'zbek tili materiallari bo'yicha kompyuter tilshunosligiga oid dastlabki tadqiqotlar olib borgan olimlar H. Arziqulov, S.Rizaev va S.Muhamedovlardir. Mazkur olimlar ham asosan kompyuter lingvistikasining statistik tahlil, avtomatik tahrir yo'nalishlari bo'yicha ishlarni amalga oshirgan.

Qoraqalpog' tili bo'yicha esa, 1979-1989-yillar oralig'ida «Sovet Qoraqalpog'istonı» gazetasi, «Amiwdarya» jurnali, Berdaq, Ajiniyoz badiiy asarlari, T. Qayıpbergenovtın «Maman biy» afsonasi, «Alpamis», «Sharyar», «Qirg'iz» dostonlari va I. Yusupovning «Tanlangan» badiiy asarlari matnlarining alfavitli, chappa alfavitli chastotali lug'atlari tuzildi. Ularning umumiy hajmi yarim million so'zdan iborat bo'ldi. Toshkent shahrida o'tkazilgan uchunchi xalqaro tyurkologlar konferentsiyasida bu xususida ma'lumotlar berildi. O'sha davrda gumanitar fanlarning bu mashinalardan foydalanish imkoniyati cheklangan edilar. Buning asosiy sababi, EHM lari faqatgina jamoa guruhlarining foydalanuvchilariga gina mo'ljallangan edi. Bundan boshqa tilshunoslarga bu mashinalardan foydalanish vaqti juda kam beriladigan edi. Albatta bunday vaziyatlar tilshunoslikda modellashirish metodlarining rivojlanish yo'lida ancha to'siqlarni yuzaga keltirdi.

Hozirgi davrda esa bunday cheklovlar mavjud emas. Imkoniyatlar ko'p. Turkiy tillarning milliy korpusini yaratish bo'yicha ishlari alla qachon boshab yuborilgan va jadal tarzda davom etmoqda. Bugungi kunda korpus lingistikasi chastotali lug'atlarga asoslanib analiz olib bormoqda [Abduraxmanova, 2021, 288].

Lekin qoraqalpoq tilshunosligida mustaqillikdan so'ngi badiiy adabiyotlar, publitsistik materiallar, matnlar asosida chastotali lug'atlarni tuzish, shu asosda qoraqalpoq tilining milliy korpusini yaratish masalalari dolzarbligicha qolmoqda.

Foydalanilgan adabiyotlar:

1. Abduraxmanova N.Z. Kompyuter lingvistikasi. Darslik. Toshkent, "Nodirabegim", 2021.
2. Айымбетов М. Лингвистикалық информатиканың семиотикалық тийкарлары. Нөкис, Илим, 2013.
3. Насыров Д.С., Айымбетов М. Бердақ таңламалы шығармаларының алфавитли жийилик сөзлиги. Нөкис, «Қарақалпақстан», 1983.
4. Rahimov A. Kompyuter lingvistikasi asoslari. (O'quv qo'llanma). Toshkent, 2011.

O'ZBEK TILI TEZAVURSI UCHUN OT SO'Z TURKUMI DOIRASIDAGI AYRIM KONSEPTLARNING PSIXOLINGVISTIK TADQIQI

Xoshimjonova Qizlarxon

NamDUning lingvistika (o'zbek tili) fakulteti magistranti

Annotatsiya. Ushbu maqolada psixolingvistikada o'zbek tili tezavursi uchun ayrim konseptlarning o'ziga xosligi va bugungi kundagi xususiyati haqida so'z boradi. Uning matnini idrok etish, tushunish jarayonida tutgan o'rni o'zbek gender lug'atini psixolingvistikada yetarlicha o'rganilmaganligi, bu esa, o'z navbatida, matn tahlili va semantik idrok etish qoidalarini aniqlash masalasida jiddiy e'tibor qaratishni talab qiladi, shuningdek, o'zbek tili tezavursi uchun mavhum otlarning psixolingvistik tahlili olib boriladi.

Kalit so'zlar: matn, retsiyent, intellektual til, lisoniy qobiliyat, affekt til.

Аннотация: В данной статье говорится об уникальности некоторых понятий для тезауруса узбекского языка в психолингвистике и их характеристике на сегодняшний день. Роль текста в процессе восприятия и понимания заключается в том, что узбекский гендерный словарь недостаточно изучен в психолингвистике, что в свою очередь требует серьезного внимания с целью определения правил анализа текста и смыслового восприятия.

Ключевое слово: Текст, получатель, интеллектуальный язык, лингвистическая, способность, аффективный язык.

Abstract: This article talks about the uniqueness and necessity of some concepts for Uzbek language thesauri in psycholinguistics. The role of gender vocabulary in the process of perceiving and understanding the text has not yet been sufficiently studied in psycholinguistics and, on the other hand, it is serious about determining the rules of text analysis and semantic perception requires attention, a psycholinguistic analysis of abstract nouns for the thesaurus of the Uzbek language is carried out.

Keywords: text, recipient, intellectual language, linguistic ability, affective language, gender vocabulary.

Tilshunos tilni kishilarning aloqa qilishi va fikrlash vositasi ekanligini e'lon qiladi, biroq tilni tadqiq etishda til egalarining, umuman, keragi bo'lmaydi. Til strukturasi uni yaratgan, undan foydalanayotgan inson faoliyatidan mustaqil ravishda tasvirlanadi. Boshqacha qilib aytganda, zamonaviy tilshunoslikning ziddiyati shundaki, inson tilini o'rganayotgan tilshunoslik insonsiz ish ko'radi...degan edi psixolingvist L.V.Saxarniy. O'z davrida mashhur psixolog, nutq bo'yicha mutaxassis N.I. Jinkin shunday yozgan edi: Agar lingvistika va psixologiya aniq tafakkur jarayonini tadqiq etishni rad etsa, ular juda katta yo'qotishga, ya'ni inson va uning nutqi tushib qolgan yo'qotishga uchraydi [Shoira Usmonova.7]. Biz yuqoridagi fikrlarga tayangan holda quyidagi izlanishlarimizga antroposentrik jihatdan yondashamiz. Psixolingvistikaning diqqat markazida inson turadi. Uning eng muhim xususiyati inson omili hisoblanadi. Binobarin, inson nutqni yaratadi va uni idrok qiladi. Psixolingvistika bu omilni nafaqat nutqning, balki eng muhimi - tilning o'zini ham xarakterini, mohiyatini aniqlaydigan majburiy omil sifatida o'z modellarining tavsiflari, asos qilib olingan hukmlariga faol tatbiq qiladi. Bunda qandaydir mavhum "umuman

odam” emas, balki real harakatdagi xotiraga ega real odamlar, ularning yoshiga oid xususiyatlari, shaxsiy tajribasi, ijtimoiy roli va h.k. nazarda tutiladi [Sh Usmonova.8.]. Aynan biz ko‘zlagan maqsad ham inson omili uchundir. O‘zbek tili tezaurusidagi konseptlarning psixolingvistik tadqiqini amalga oshirishda abstrakt otlardan foydalanishni qaror qildik. Negaki, inson hayotida ma’lum bir his-tuyg‘u, anglab bo‘lmas mavhum tushunchalar qurshovi oldida qolganida abstrakt otlardan foydalanadi. Inson o‘z psixologiyasini so‘zlar orqali namoyon etadi. Ma’lumki, Shveytsariyalik tilshunos Sharl Ballining ta’limotida hayajonli nutq holatini o‘rganish masalasiga alohida e’tibor qaratilgan. Olimning fikricha, tilda affektiv (hayajonli) vaziyatning yuzaga kelishi ikki asosiy maqsadni ko‘zlaydi: 1) so‘zlovchining his-tuyg‘usi, kayfiyatini ifodalash hamda ma’lum qilinayotgan axborotga baho berish; 2) muloqot ishtirokchilariga ta’sir o‘tkazish maqsadida ma’lum turdagi lisoniy vositalardan foydalanish ... Sh.Balli lisonda intellektual va affektiv xususiyatlar munosabatini aniqlashda psixologiya fanida ushbu xususiyatlarning inson ongida alohida aks etishi haqidagi g‘oyaga asoslanadi. Tafakkur faoliyati hamda ongda bunday ikki tomonlamalik to‘g‘ridan to‘g‘ri til tizimiga ham ko‘chadi va Balli “intellektual” va “affekt” tillarini farqlaydi. Tilning hayajonli- affektiv xususiyatlari haqidagi Ballining fikrlari ekpressivlikning nutqiy faoliyatdagi o‘rnini aniqlash uchun muhim [Nasirova G.2022. B.210]. Intellektual til – bu lot. aql, idrok, bilish tushunish, affekt til – bu (lot. ruhiy hayajon) tashqi yoki ichki ta’sirlar tufayli tez paydo bo‘lib, qisqa muddatli aksariyat hollarda “portlash tarzida bo‘ladigan kuchli hissiyot holatidir. [Izohli lug‘at.@dasturlar2018]. Demak, ichki va tashqi nutq orasida inson ma’lum tafakkur tarzini qo‘llaydi. Yuzlab yaqin ma’nodosh so‘zlar orasidan vaziyat uchun munosibini tanlash aynan insonning tafakkuriga yuklatilgan vazifadir.V.fon. Gumboldt aytganidek, nutqiy faoliyat negizida “lisoniy qobiliyat” yotadi[Shoira Usmonova.13]. Ya’ni, ichki nutq – lisoniy qobiliyat (so‘z tanlay olish) – tashqi nutq sxemasi holatida qarasak, biz psixolingvistik jihatdan asosiy urg‘uni lisoniy qobiliyat markaziga berishimiz zarur, negaki aynan shu markazda analiz va sintez jarayonlari amalga oshadi va so‘zlovchi semalar ichidan vaziyat uchun kerakli so‘zni tanlay olishi zarurdir. Insonning lisoniy qobiliyati uning ichki dunyosi, uni o‘rab turgan muhit bilan ham, undagi insonlar bilan ham uzviy bog‘liqdir.

1980-yillarning o‘rtalarida bir qancha tilshunos va psixologlar Priniston universitetida leksik-semantik ma’lumotlar bazasini yarata boshladilar. [Alessandro Agostini, Timur Usmanov, Ulugbek Khamdamov, Nilufar Abdurakhmonova, and Mukhammadsaid Mamasaidov. 2021]. Bu esa WordNet qidiruv tizimini yaratilishiga zamin bo‘ldi. Bunda, asosan, sinonimlar hamda giponimlarga asoslandi. Biz yuqorida til va nutq oralig‘idagi insonning “lisoniy qobiliyat”i tushunchasini psixolingvistikada retsipyent muhim vazifalardan biri ekanligini ko‘rib o‘tdik. Uni yanada rivojlantiruvchi, shakllantiruvchi, retsipyentga yanada qulayroq bo‘lishini ta’minlash maqsadida, tezaurus lug‘atlardan foydalanish usulini taklif qilamiz. Negaki, bu turdagi lug‘atlardan foydalanishning bir talay afzalliklari va qulayliklari mavjud. Ushbu lug‘atning o‘ziga xos xususiyatlari: lug‘atdan har qanday inson onlayn foydalanishi mumkinligi; lug‘atda kamida uch til mavjud bo‘lib, foydalanish imkoniyati keng ekanligi; qidiruv tizimining mavjudligi; qidirilayotgan so‘zning izohi, sinonimi, giperonim, giponimi, antonimi, qaysi konseptga kirishi batafsil berilganligi; har bir so‘zning o‘qilishi(audio) ham kiritilganligi va bir qancha ma’lumotlar mavjudligi, nafaqat to‘g‘ri yozish bir vaqtning o‘zida har bir berilayotgan so‘z haqida har tomonlama to‘liq ma’lumotga ega bo‘lishini ta’minlaydi.

Biz misol tariqasida **farishta** so‘zini olamiz:

1.Morfologiya: O‘zak farishta— tub so‘z . Ot so‘z turkumiga mansub, turlanadi.

2. Talaffuz qilinishi: birlikda [farishta]

3. Ma'no tarkibi: islom dinida va boshqa dinlarda yaratganning amrini bajaruvchi mavjudot, ko'chma ma'nolari: yaxshi xulq-atvorga ega bo'lgan va go'zal ayollarga, pokiza, musaffo, begunoh ma'nolarini anglatadi.

A) sinonimlari: "malak", "hur", "g'ilmon"

B) antonim: farishtasiz -kishini o'ziga tortmaydigan, ko'rksiz, yoqimsiz, farishtasiz odam;

D) giperonim (umumiy): mavjudodlarning turi

E) giponim(xususiy): ezgulik farishtasi, yovuzlik farishtasi.

F) meronim (qism): farishtali – fayzli, serbaraka, yoqimli, halol, pokiza

4. Frazologizmlarda yoki turg'un birikma shaklida qo'llanishi: ayolning sarishtasi – ro'zg'orning farishtasi(maqol). Sarishtali uy – farishtali uy(maqol). Farishtali odam. Yuzining farishtasi.

5. Boshqa tillarda qo'llanishi. Ruscha – ангел, ingliz – angel, fransuzcha – ange.

Nutqni idrok qilish til elementlarini eshitish yoki ko'rish resepsiyasi (qabul qilish), ularning o'zaro aloqasini aniqlash va ularning mazmuni haqidagi tasavvurlarning shakllanishini o'z ichiga oladi. Shunday qilib, nutqni idrok qilish ikki bosqichda - nutqiy ifodani idrok qilish va tushunish orqali amalga oshadi. Psixologiyada tushunish bevosita idrok qilinadigan nutq (tovushlar) oqimi ortida turgan umumiy mazmunni ochish sifatida talqin qilinadi. Bu idrok qilinadigan nutq orqasida turgan ma'noning haqiqiy mazmunga aylanish jarayonidir[Usmonova: 59]. Masalan, "Farishta" ifodasining ma'nosi qanday "nutqsiz kontekstda" va kim tomonidan aytilganiga ko'ra turlicha bo'lishi mumkin. Olib borilgan psixologik tadqiqotlarimiz natijasida quyidagilarni havola etamiz. Agar bu diniy ulamolar tomonidan aytilayotgan bo'lsa bosh ma'nosida qo'llaniladi va jannatdagi malak, hurlar nazarda tutiladi. Agar mazkur ifoda "mening farishtam" birikmasi tarzida bo'lsa, 12 yoshli bola tilidan- "Onasi" nazarda tutiladi, 20- 25 yoshlar oralig'idagi yigit tilidan aytilsa "bokira qiz" tushuniladi, 35-40 yoshlar oralig'idagi inson tomonidan aytilganda uning "qizalog"iga ishora qilinadi. Yoki begunoh ayblangan insonga nisbatan ham qo'llanilishi mumkin.

Ma'lumotlarni qidirishda veb-hujjatlarni avtomatik ravishda yuklab oladigan veb- brauzer dasturlari mavjud. Qidiruv ilovalari turli sohalarida, masalan, moliya, tibbiyot, yangiliklar va hokazolarda qo'llanilishi mumkin [Abdurakhmonova. N., Ismoilov.A and G.Toirova.]. Psixolingvistikada gender ma'nolarni ifodalashda foydalaniladigan lug'atlar mavjud, ular: gender assotsiativ maydonlari lug'atiga kirishdan oldin beriladi, gender lug'atning lug'at yozuvi jadval shaklida bo'ladi. [A.B.Рудакова.2018:27.], masalan :

Izohli lug'atda: Malak, hur, g'ilmon ma'nolarida keladi. Islom dinida yaratganning amrini bajaruvchi g'ayritabiiy mavjudot.	Psixolingvistik tahlil	
	Erkaklar uchun: (25-30yosh)	Ayollar uchun:(25-30 yosh oralig'i)
	Begunoh inson, ibodat qiluvchi, ona,(jufti halol)ayol, bokira qiz,go'dak, chaqaloq, itoatli bo'lish, shirin so'zlik, go'zallik, jozibalilik, opa, singil, xushbichim, xushro'y...	Ibodot qiluvchi, ona, qizaloq, chaqaloq, opa - singil, dugona, chiroyli, xushbichim,

Ko‘rinib turibdiki, ayollarga nisbatan erkaklar tomonidan **farishta** so‘zining xususiyati va semalarining darajasi yuqori. Ammo ayollarda uchragan dugona semasi orqali erkaklar semasi bilan farq qiladi. Demak, **farishta** so‘zining sof ayollik semasi – bu dugona. Farishta so‘zining sof erkaklik semasi – bu (jufti halol)ayoliqa nisbatan semasi hisoblanib bunda intilish, yaqinlashish, havas qilish kabi xislatlar bilan bu izlanishimiz o‘zini oqlaydi. Farishta konseptini tezaurus yaratayotganda mana shu semalarni ham qo‘shib ifodalasak, o‘ylaymizki, yanada mukammal ko‘rinishga keladi. Farishta so‘zi makrostrukturani tashkil etuvchi sema hisoblangan diniy konseptlar: namoz, masjid, ro‘za, haj, savob, gunoh kabi tushunchalar qatoridagi sistemaning bir bo‘lagi hisoblanadi. [Tohir Malikning Shaytanat 1-kitob(1989)]ida *...bir paytlar bu ajdaho ko‘ziga farishta bo‘lib ko‘ringan edi*. Yoki [Shaytanatning 2-kitob (1991)]ida *Oyim aytdilar, eringizga ertalabdan jilmaysangiz, uyingizga farishta kiradi, dedilar,tarzida yoki ...ko‘rgan qalloblar oldida seniki farishtaning o‘zi-ku? – ol-a! sen shu farishta bilan o‘n yil emas o‘n kungina birga bo‘lsang bilarding*, qabilidagi gaplarda ham farishta so‘zining turli semalarini yaqqol ko‘rishimiz mumkin. Agar ushbu so‘zning ayollik semasida ko‘rsak ayollik konseptini tashkil etuvchi yana bir mikrosistema vujudga keladi. Sinonimlari: *xotin, qiz*. Antonimi *erkak, yigit*. Birikmalar: *ayol rashki, ayol muhabbati, ayol muhabbati, ayoljozibasi, musulmon ayol, bardoshli ayol, insonparvar ayol, ayol salomatligi, ayol qalbi, mushtipar ayol* kabi bir necha birikmalarni o‘zbek tili tezaurusida uchratamiz.

L.V.Sherba nutqiy faoliyat va til materiali o‘rtasidagi munosabatni quyidagicha belgilaydi: “Nutqiy faoliyat til materialini yaratadi. Til tizimi tilshunoslar tomonidan til materialidan chiqarib yuboriladi. Nutqiy faoliyat ham til materialini ishlab chiqaradi, ham til tizimiga o‘zgartish kiritadi. Mana shunday hodisaning yana biri “ijtimoiy guruh” hayotining u yoki bu davridagi muayyan sharoitda so‘zlashuv va tushunishlar majmuyi” hisoblanadi [Usmonova Sh.]. Masalan, **farishta** so‘zini aks holda qo‘llash ya’ni “nopok yo‘ldagi inson” brikmasi muayyan davrdagi tushunchalar majmuyiga misol bo‘la oladi. Psixolingvistika tilning boshqa xususiyatlarini, xususan, til birliklaridan nutqiy jarayonda foydalanish va ularda shaxsning xususiyatlari, ruhiyati va tafakkuri, xarakteri va kayfiyati bilan bog‘liq ma’no, tushuncha ifodalanishi kabi masalalarni o‘rganadi. Til va millat, til va xalq, til va shaxs, til va tafakkur, til va jamiyat kabi tushunchalar bir-biridan ajralmas, bir-biriga ma’no jihatdan bog‘langan bo‘lish bilan birlikda, bir-birini boyitib, to‘ldirib boruvchi hodisalardir. Kishilar til yordamida fikr almashadilar. Til tafakkur bilan chambarchas bog‘liq bo‘lib, fikrni moddiylashtiradigan, reallashtiradigan vositadir. Tafakkur inson oliy nerv sistemasining faoliyati bo‘lib, tilda o‘z ifodasini topadi. Uning moddiy qobig‘i tildir. Demak, til materiallarisiz fikr yuzaga chiqmaydi. Til bilan tafakkur murakkab qarama-qarshiliqlarga boy bo‘lgan dialektik birlikni tashkil etadi. Ko‘rinadiki, til jamiyatning asosiy aloqaaralashuv vositasi sifatida tushuniladi, shuning uchun tilni har tomonlama o‘rganish va tahlil qilish lozim. Qolaversa, matnning tuzilishi va semantik idrok etish mezonlarini aniqlash, birinchi navbatda, nutq faoliyatining operativ tarkibini psixolingvistik tahlil qilish asosida matn strukturasi tarkibiy qismlarini o‘rganish bilan bog‘liq bo‘lib, u hal qiluvchi ahamiyatga ega.

Tezauruslaridagi konseptlarning psixolingvistik tadqiqi, avvalo, ma’lum xalqning, jamoaning tushunchasi, madaniyati bilan bog‘liq bo‘lib, bu bevosita milliylikni o‘zida aks ettiradi. Zero, akademik L.V.Sherba ta’kidlaganidek: “Milliy tillarning lug‘atlari haqiqiy xazina, milliy tushunchalar, millat hayoti va aqlining o‘ziga xos qomusidir”. Yevropada ideografik lug‘atlarga qiziqish uyg‘otgan L.M. Roje o‘z lug‘atining so‘z boshisida shunday yozadi: “Biz aytmoqchi

bo'lgan fikrimizni qanchalik aniq tasavvur qilmaylik, ba'zan biz shunday bir vaziyatga tushib qolamizki, bizga so'z yetishmay qoladi. O'z fikrimizni aniq ifodalash uchun bitta so'z o'rniga bir necha ijobiy va salbiy, kuchli va kuchsiz, ba'zan esa keraksiz so'zlarni ishlatib yuboramiz. Bu lug'atning eng muhim xususiyati shundaki, bir fikrning barcha qirralarini ochib beruvchi so'z va iboralarning boy to'dasini tortiq qiladi [Ne'matov: 23–27]. So'zlarning mavzuiy to'dalar qurshovida berilishi til o'rganayotgan o'zga millat vakillariga til sohiblarining dunyoqarashi, o'sha millatning turmush tarzi, madaniyati, ma'naviyati va o'ziga xos xususiyatlari bilan tanishtirish imkoniyatini beradi.

Foydalanilgan adabiyotlar:

1. Alessandro Agostini, Timur Usmanov, Ulugbek Khamdamov, Nilufar Abdurakhmonova, and Mukhammadsaid Mamasaidov. 2021. UZWORDNET: A Lexical-Semantic Database for the Uzbek Language (<https://aclanthology.org/2021.gwc-1.2>). In Proceedings of the 11th Global Wordnet Conference, pages 8–19, University of South Africa (UNISA). Global Wordnet Association.
2. A.B. Рудакова. Психоллингвистика и лексикография. «Издательство РИТМ», 2018.
3. Izohli lug'at.@dasturlar 2018.
4. Nasirova Gulnora Madaminovna. Psixolingvistik tadqiqotlar ahamiyatining dolzarbligi. ResearchBip (14).2022.
5. N. Z. Abdurakhmonova, A. S. Ismailov and D. Mengliev, «Developing NLP Tool for Linguistic Analysis of Turkic Languages,» 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation, 2022, pp. 1790-1793, doi: 10.1109/SIBIRCON56155.2022.10017049.
6. Ne'matov X. So'z va uning til va nutqdagi o'rni. // O'zbek tili va adabiyoti. 1988. 6-son.
7. 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.
8. 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.
9. Shoiri Usmonova. Psixolingvistika. "Universitet" 2014.

BILIMLAR BAZASINI ONTOLOGIK MODELLARDA IFODALASH

Qodirova Zebo Gulboyevna

Mirzo Ulugʻbek nomidagi

Oʻzbekiston Milliy Universiteti magistranti

E-mail: qodirovazebo5450@gmail.com

Annotatsiya: Bugungi kunda oʻzbek tilshunosligida kompyuter lingvistikasiga doir qator izlanishlar olib borilmoqda, jumladan, mashina tarjimai, 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. Axborotlar oqimining koʻpayishi bilan uni saqlash, taqdim etish, rasmiylashtirish, va tizimlashtirish, shuningdek, avtomatik qayta ishlashning yangi usullarini izlash zarurati paydo boʻladi. Ontologiya esa axborotlar oʻrtasidagi semantik munosabatlarga asoslangan qidiruv tizimining joriy etilishi uchun dastlabki asos boʻlishi bilan eʼtiborlidir.

Kalit soʻzlar: Ontologiya, RDF, XML, OWL, spetsifikatsiya, konseptualizatsiya, sunʼiy intellekt, semantik veb.

Annotation: Today in Uzbek linguistics, a number of researches on computer linguistics are being carried out, including machine translation, corpus technology, computer lexicography and the development of linguistic bases of software applications, lexicographic text processing and linguostatistical analysis. With the increase in the flow of information, there is a need to search for new methods of its storage, presentation, formalization, and systematization, as well as automatic processing. Ontology is important because it is the initial basis for the introduction of a search system based on semantic relations between information.

Keywords: Ontology, RDF, XML, OWL, specification, conceptualization, artificial intelligence, semantic web.

Аннотация: На сегодняшний день в узбекской лингвистике проводится ряд исследований по компьютерной лингвистике, включая машинный перевод, корпусные технологии, компьютерную лексикографию и разработку лингвистических основ программных приложений, лексикографическую обработку текстов и лингвостатистический анализ. С увеличением потока информации возникает необходимость поиска новых способов ее хранения, представления, формализации и систематизации, а также автоматической обработки. Онтология важна, поскольку является исходной основой для внедрения поисковой системы, основанной на семантических отношениях между информацией.

Ключевые слова: онтология, RDF, XML, OWL, спецификация, концептуализация, искусственный интеллект, семантическая сеть.

Xorijda bilimlar bazasini ontologik modellarda ifodalash [Dang, J., Hedayati, A., Hampel, K., Toklu, C., 2008: 2], [Galárraga, L., Teflioudi, C., Hose, K., Suchanek, F. M. 2015: 3], [Janik, M., Kochut, K.J. 2008: 4], [Aminu, E.F., Oyefolahan, I. O., Abdullahi, M.B., Salaudeen, M.T. 2020: 5] boʻyicha tadqiqotlar talaygina. Intellektual tizimlarning oʻziga xos xususiyati - muayyan

fan sohasidagi muammolarni hal qilish uchun zarur bo'lgan bilimlarning mavjudligi. Bilim ma'lumotlarga qaraganda, murakkabroq ma'lumot kategoriyasidir. Bilim nafaqat individual faktlarni, balki ular o'rtasidagi munosabatlarni ham tasvirlaydi, shuning uchun bilim ba'zan tuzilgan ma'lumotlar deb ataladi. Bilimlarni empirik ma'lumotlarni qayta ishlash orqali olish mumkin. Ular shaxsning amaliy faoliyati natijasida orttirilgan tajribasini umumlashtirishga qaratilgan aqliy faoliyati natijasini ifodalaydi [A. A. Sharipbay, A. S. Mukanova, R.S. Niyazova, B.SH. Razakhova, A. S. Omarbekova, N. Z. Abdurakhmonova: 6]. Ontologiya bilimlarni boshqarishning murakkab usuli hisoblanadi. Ontologiyalarni loyihalashda shartli ravishda ikkita yo'nalishni ajratish mumkin, ular ma'lum vaqtgacha alohida ishlab chiqilgan. Birinchisi, ontologiyani matematik jihatdan aniq aksiomalarga asoslangan rasmiy tizim sifatida ifodalash bilan bog'liq. Ikkinchi yo'nalish hisoblash tilshunosligi va kognitiv fan doirasida ishlab chiqilgan. U yerda ontologiya deganda, faqat inson ongida mavjud bo'lgan, tabiiy tilda (yoki boshqa belgilar tizimi orqali) ifodalanishi mumkin bo'lgan mavhum tushunchalar tizimi tushunilgan. Shu bilan birga, odatda, bunday tizimning aniqligi yoki izchilligi haqida hech qanday taxminlar qilinmaydi.

Shunday qilib, ontologiyalarni yaratish va tadqiq qilishning ikkita muqobil yondashuvi mavjud. Birinchisi (rasmiy) mantiqqa asoslanadi (birinchi tartibli predikatlar, tavsiflovchi, modal va boshqalar). Ikkinchisi (lingvistik) tabiiy tilni (xususan, semantikani) o'rganishga va korpus deb ataladigan katta matn massivlarida ontologiyalarni qurishga asoslangan [A.A. Sharipbay, A. S. Mukanova, R.S. Niyazova, B.SH. Razakhova, A. S. Omarbekova, N. Z. Abdurakhmonova: 6].

Ontologiyani ishlab chiqish jarayoni 4 bosqichni o'z ichiga oladi.

1. Spetsifikatsiya - ontologiyalarni yaratish maqsadlariga, uning maqsadli qo'llanilishiga va potensial foydalanuvchilarga bo'ysunadi. Ma'lumotlar bazasiga kiritilgan elementlar ro'yxatini o'z ichiga olgan loyiha shakli, dizayni deyish mumkin.
2. Konseptualizatsiya - aniq modelda fan bilimlarini tuzilishini ta'minlaydi. Konseptualizatsiya – tushunchalar to'plamini aniqlash jarayonini belgilab beruvchi yo'l demakdir.
3. Rasmiylashtirish - Rasmiylashtirish bu formallashtirishdir. Obyekt yoki hodisaning mohiyatini ko'rsatishda belgilangan shartlarni qayta ishlash jarayoni. Tanlangan dastur tilida ba'zi bir algoritim yordamida ma'lum modelni yaratish bosqichi.
4. Amalga oshirish - amalga oshirish bosqichida ontologiyaga mos keladigan va bilimlarni ifodalash tillaridan birida ontologiya modeli yaratiladi.

Konseptualizatsiya ikki bosqichni o'z ichiga oladi, ular doirasida muammolar hal qilinadi [Hazarov D.A. 2003: 7]:

- 1) atamalar lug'atini yaratish;
- 2) konseptlarning tasniflash iyerarxiyalarini qurish.

Konseptualizatsiyaning tabiati uning qanday paydo bo'lganiga yoki qanday yaratilganiga bog'liq. Konseptualizatsiya – bu o'z tajribasidan tegishli ma'lumotlarni olish va umumlashtirishga olib keladigan jarayon. Konseptualizatsiya – bu muayyan ma'lumotlarning o'zi va muayyan vaziyatlardan, vakillik tillaridan mustaqildir. Ikkinchi operatsiya spetsifikatsiya, chunki ontologiya vakillik tilida konseptual izlanishni belgilaydi. Murakkablik va aniqlik darajasidan tashqari, hal qiluvchi narsa shundaki, ontologiyalar konseptualizatsiyaning tilga bog'liq spetsifikatsiyalari sifatida, muloqotning asosi bo'lib, umumiy tushuncha o'rnatiladigan ko'prikdir [Huang C.R., Calzolari N., Gangemi A., Lenci A., Oltramari A. 2010: 8].

Bugungi kunda mavjud bo'lgan eng mashhur OWL va RDF muharriri Protégé texnologiyasi hisoblanadi. Muharrirni dastur veb-saytidan bepul yuklab olish mumkin. Dastur Java tilida amalga oshiriladi va shuning uchun Java dasturini oldindan dasturchi kompyuteriga o'rnatish talab etiladi. Odatda, Java mashinasi Protégé bilan birga foydalanuvchi kompyuteriga o'rnatilishining bir qismi sifatida qaraladi. Protégé OWL ontologiyalarini yaratish uchun qulay muharrir bo'lib, uning imkoniyatlari ham bir qadar kengroq.

Bilamizki, World Wide Web – bu online tizim orqali kirish mumkin bo'lgan giperbog'langan hujjatlar tizimi. Bu oddiy hujjatdan farq qiladi, chunki u nafaqat matnni, balki boshqa ma'lumotlarni ham o'z ichiga olishi mumkin: rasmlar, musiqa va video fayllar va boshqalar. Bunday hujjatlarni maxsus dasturlar yordamida ko'rish mumkin. Olimlarning fikrlariga ko'ra, World Wide Webning vorisi sifatida sun'iy intellekt yoki semantik veb deya e'tirof etilmoqda. Semantik veb butunjahon online tarmoq tizimining avlodi bo'lib, yuqoridagilarga qo'shimcha sifatida ma'lumotlar bazasi semantikasi tavsiflarini taqdim eta olish imkoniyatiga ega hisoblanadi. Bunday tavsiflar nafaqat foydalanuvchilar uchun balki, dasturiy ta'minot uchun ham tushunarli bo'ladi. Aqlli tizimlar tomonidan ma'lumotlar bazasi mazmunini tushunish online tizimda ma'lumotni aniqroq qidirish imkonini beradi, qolaversa, bu ma'lumotlarni avtomatik tarzda tartibga solish hamda boshqa dasturiy ta'minotlar bilan almashinish imkonini beradi.

Veb-ma'lumotlar semantikasini tavsiflash uchun RDF (Resurs tavsifi ramkasi) tilidan foydalaniladi. Bu til vebhujjatlarning resurslarini tavsiflash uchun mo'ljallangan, ammo bu to'liq ontologiyalarni ifodalash uchun yetarli emas. Shu sababli, Web ontologiyasini tavsiflash uchun maxsus til, OWL tili ham taklif qilindi. Ushbu til RDF tiliga mos keladi, ya'ni har bir OWL hujjati ham RDF hujjatidir. RDF XML sintaksisi asosida qurilgan bo'lib, u Internetda ma'lumotlar almashinuvi uchun standart sifatida ishlatiladi. Shuning uchun har bir OWL hujjati ham haqiqiy XML hujjatidir [Eshonqulov H. 2022: 9]. XML tilining asosiy elementi bu **teglar** deb nomlanuvchi hujjatni belgilash uchun bevosita foydalaniladigan belgilar ketma-ketligi hisoblanadi. Teg burchakli qavs ichiga olingan nomdir, masalan: <name>. Boshlang'ich va yakunlovchi teglar bir xil nom bilan nomlanadi, biroq ular urtasidagi farq yakunlovchi tegning burchak qavs ichida nom oldidan </name> belgi qo'yish orqali farqlanadi. Dastlab XML hujjatlarni belgilash uchun yaratilgan edi, biroq hozirda bu tildan boshqa tarzda foydalaniladi. XML ma'lumotlari hozirda bir kompyuterdan ikkinchisiga o'tkazilishi kerak bo'lgan ma'lumotlarning matnli ko'rinishi sifatida keng qo'llanilmoqda. Boshqacha qilib aytganda, agar ilgari XML kompyuterlar uchun tushunarli bo'lgan hujjatlarni belgilash uchun ishlatilgan bo'lsa, hozirda XML allaqachon tuzilgan ma'lumotlarni ketma-ket (matn) shaklga aylantirish uchun ishlatiladi.

OWL veb-ontologiya tili shunchaki ma'lumot berib qolmasdan, ma'lumotlar mazmunini qayta ishlashga muhtoj bo'lgan ilovalarda foydalanish uchun mo'ljallangan. OWL XML, RDF tillaridan ko'ra mashina talqinini osonlashtiradi. RDF (Resource Description Framework) bu online tizimda joylashgan ma'lumotlarni tavsiflash imkonini beruvchi tildir. RDF tavsif tili bo'lish bilan birgalikda shu tilda yozilgan ma'lumotlarni to'liq o'qish va almashtirishni ta'minlaydigan turli xil qo'shimcha dasturiy modellar ma'nosida ham qo'llaniladi.

Xulosa. Bugungi kunda zamonaviy axborot-texnologiyalari inson hayotiy faoliyatining asosiy shakliga aylanib ulgurdi. Shuning uchun ham yangi termin va atamalarning kirib kelish jarayoni jadallashdi. Tushunchalarning ko'payishi axborotdan foydalanish jarayoniga salbiy ta'sir etishi mumkin, shu jihatdan, ontologik bilimlar bazasi matnli ma'lumotlarning ma'nosi bo'yicha

samarali qidiruvni tashkil qilish, fikrlash mexanizmlarini amalga oshiradigan tizimlarni yaratish, turli xil intellektual bilimlar uchun umumiy bilim bazalarini qurish va ulardan foydalanish tizimlarini yaratishda ahamiyati beqiyosdir.

Foydalanilgan adabiyotlar:

1. Лутошкина, Н. В. Модели знаний и онтологии : учеб. пособие / Н. В. Лутошкина ; СибГУ им. М. Ф. Решетнева. – Красноярск, 2021. – 80 с.
2. Dang, J., Hedayati, A., Hampel, K., & Toklu, C. (2008). An ontological knowledge framework for adaptive medical workflow. *Journal of biomedical informatics*, 41(5), 829-836.
3. Galárraga, L., Teflioudi, C., Hose, K., & Suchanek, F. M. (2015). Fast rule mining in ontological knowledge bases with AMIE $\$ \$ + \$ \$ +$. *The VLDB Journal*, 24(6), 707-730.
4. Janik, M., & Kochut, K. J. (2008, August). Wikipedia in action: Ontological knowledge in text categorization. In *2008 IEEE International Conference on Semantic Computing* (pp. 268-275). IEEE.
5. Aminu, E. F., Oyefolahan, I. O., Abdullahi, M. B., & Salaudeen, M. T. (2020). A review on ontology development methodologies for developing ontological knowledge representation systems for various domains.
6. A. A. Sharipbay, A. S. Mukanova, R.S. Niyazova, B.SH. Razakhova, A. S. Omarbekova, N. Z. Abdurakhmonova ONTOLOGIYA, SEMANTIK TEXNOLOGIYALAR. Elektron darslik. Nur-Sulton. 2021 yil.
7. Назаров Д.А. Интеллектуальные информационные технологии. Методы представления знаний. Редакция: октябрь 2013
8. Huang, C. R., Calzolari, N., Gangemi, A., Lenci, A., & Oltramari, A. (Eds.). (2010). *Ontology and the lexicon: A natural language processing perspective*. Cambridge University Press.
9. Eshonqulov Hakim Ilhomovich Ontologiyalar va bilimlarni ifodalash, avtomatlashtirilgan axborot tizimlarining ontologiyasini qo'llash asoslari [Matn] / H.I. Eshonqulov. Buxoro: «Sadridin Salim Buxoriy» Durдона, 2022. -128 b.
10. Abduraxmonova, N. Z. «Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aforef.» (2018).
11. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. *Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta*. 2016;2 (38):12-7.
12. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
13. 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.

14. 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;5(03):89-100.
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. 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
16. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019.

O‘ZBEK TILI VALENTLIK LUG‘ATINI YARATISHDA KORPUSGA ASOSLANGAN TAHLIL

Latipova Gulasal

Toshkent davlat o‘zbek tili
va adabiyoti universiteti tayanch doktoranti

Annotatsiya: Ushbu tezisdagi korpusga asoslangan valentlik lug‘atini yaratishga doir mulohazalar tahlilga tortilgan. Valentlik lug‘atini yaratish kompyuter leksikografiyasi uchun muhim lingvistik resurs sifatida parallel korpuslarda tarjima muqobillarini aniqlash, so‘z birikmasi va barqaror ifodalarni farqlashda muhim ahamiyat kasb etadi. O‘zbek tili elektron korpusi tahlili yordamida valentlik lug‘atini shakllantirishda korpusning kontentini kengaytirish va janrlar xilma-xilligiga e’tibor berish lozimligi qayd etilgan.

Kalit so‘zlar: valentlik lug‘at, korpusga asoslangan tahlil, kompyuter leksikografiyasi.

Annotation: In this thesis, considerations on the creation of a vocabulary of valence based on the corpus are analyzed. Building a valence dictionary as an important linguistic resource for computer lexicography is important for identifying translation alternatives in parallel corpora, for distinguishing word combinations and stable expressions. With the help of the analysis of the electronic corpus of the Uzbek language, it was noted that it is necessary to expand the content of the corpus and pay attention to the variety of genres in the formation of the valence dictionary.

Keywords: valency dictionary, corpus-based analysis, computer lexicography.

Korpus lingvistik tadqiqotlar uchun muhim o‘rganish vositasi hisoblanadi. Uning yordamida tilning nutq aktlarining turli vaziyatlarda voqelanishi aks etadi. Bugungi kunda korpus texnologiyasi rivojlanayotgan soha sifatida lingvistikaning turli yo‘nalishlarida o‘zining amaliy jihatlari bilan ma’lumotlarni o‘rganish va tahlil qilish uchun muhim manba bo‘lib xizmat qilmoqda [Abduraxmonova N.: 2016, 2017, 2018, 2019].

Leksikografik tadqiqotlarning aksariyati (ayniqsa, zamonaviy tilshinoslik)da korpusga asoslangan tahlil asosida lug‘at yaratish borasida tajribalar amalga oshirilib, ijobiy natijalarga erishilgan. Masalan, ingliz tili doirasida yaratilgan bir tili izohli lug‘atlarni olsak, ularning strukturasi va kontentini shakllantirishda korpus tahlili natijasida olingan namunalar tilning leksik imkoniyatini ochib beish uchun katta ahamiyatga ega bo‘lgan.

Tarjima lug‘atlar misolida shuni qayd etish o‘rinli, parallel korpuslar yordamida u yoki bu tushunchani ifodalashda, ayniqsa tarjima ekvivalentlarini berishda korpus juda katta hamaiyat kasb etadi. Manbalarda ekvivalentning statistik ekvivalent, tarjima ekvivalentligi, tizim ekvivalentligi, semantik ekvivalentlik, qoida ekvivalentligi, substantive ekvivalentlik hamda pragmatik ekvivalentlik kabi turlarga ajaratiladi [Krzyszowski (1990)].

Lug‘atlarni shakllantirishda korpus tahlili katta rol o‘ynaydi:

- Tarjima birliklari va ekvivalentlar;
- Valentlik lug‘ati
- Grammatik lug‘at
- Chastotali lug‘at
- O‘quv lug‘at

- Idiomatik lugʻat
- Izohli lugʻat
- Assotsiativ lugʻat
- Tezaurus yoki ideografik lugʻat

Grammatik va lugʻatshunoslikda valentlik juda muhim sanaladi. Valentlik masalasi, shuningdek til taʼlimi uchun ham ahamiyli sanaladi. Chunki u, ayniqsa til oʻrganuvchilari uchun muhim rol oʻynaydi. Valentlik lugʻat lingvistik tadqiqotlar, amaliy lingvistika, grammatika, leksikografiya sohalari uchun asosiy tadqiqot masalalari uchun yechim boʻlishi mumkin.

Tomas Herbst [Walter de Gruyter, 2004] valentlik va leksikografiya oʻrtasidagi munosabatni fransuz lingvisti Lusien Tesniere tomonidan asos solingan tobelik nazariyasi bilan bogʻlaydi. L.Tesnier [Lucien Tesniere: 2015] valentlik masalasini feʼl soʻz turkumi doirasida oʻtimli feʼl misolida tushuntirar ekan unng toʻrt kategoriyasi mavjudligini qayd etadi: avalant (*It rained*), monovalent (*Alfred slept*), divalent (*Alfred saw Bernard*), trivalent (*Alfred gave Bernard a pen*). Demak, olim sintaktik valentlik misolida soʻzlarning oʻzaro kombinatsiyasi uchun, yaʼni soʻzlarning oʻzaro grammatik va semantik jihatdan bir-biriga bogʻlanishini misol sifatida keltirib oʻtadi. T. Herbstning fikricha, ushbu konsept keyichalik Yevropaning boshqa davlatlarida rivojlantirilgan. Ingliz tili boʻyicha Emons (1974, 1978) tomonidan ingliz tili feʼl soʻz turkumi uchun valentlik modeli ilk bor shakllantirilgan degan qarashni olim inkor etmaydi.

Korpus tahlili orqali valentlik lugʻatini yaratish oʻzbek tili oʻquv lugʻati uchun dolzarb masalalardan biridir. Korpusning hajmi lugʻatda aks etuvchi maqolalarning umumiy hajmi va sifatiga taʼsir etadi. Bu narsa korpus tahlili orqali olinayotgan nafaqat chastota balki soʻzlar qurshovida aks etgan semantik va grammatik bogʻlanish orqali ularning strukturasi hamda lingvistik modeli haqida tasavvur hosil qilishga yordam beradi.

Bizning tadqiqotimiz uchun oʻzbek tili elektron korpusi yaratilgan boʻlib, unda turli janrga tegishli hamda bir nechta uslub doirasidagi matnlar orqali qidirilayotgan soʻzning leskik maʼnosini berishda foydalanish mumkin.

Masalan, oʻzbek tili uchun *bermoq* feʼli orqali fikrimizni asoslashga harakat qilamiz. Oʻzbek tilining izohli lugʻatida quyidagicha keltirilgan:

1. ***Qoʻliga tutqizmoq; qoʻliga yoki ixtiyoriga topshirmoq.*** (Berganning betiga qarama. Maqol)

2. ***Ehtiyojini, talabini qondirmoq; yoʻq narsa bilan taʼminlamoq.***
(Ekinga suv bermoq).

3. ***Baxsh etmoq, bagʻishlamoq; ruhiy holat, kayfiyat, koʻrinish paydo qilmoq.***
Musiq dillarga orom beradi

4. ***Biror muassasaga oʻrnashtirmoq, kiritmoq.*** (Bolani bogʻchaga bermoq).

5. ***Mahsulot yetishtirib, tayyorlab yetkazib turmoq; oʻz mahsuli bilan taʼminlamoq.***
(Bu xalq xoʻjaligiga nima beradi?)

6. ***Nashrga kiritmoq, bosmoq, eʼlon qilmoq; eshittirmoq; koʻrsatmoq.***
(Maqolangizni kelgusi sonda beramiz.)

7. ***Ayrim otlar bilan qoʻshma feʼl yasaydi:*** (ahamiyat bermoq, bardosh bermoq, javob bermoq, boy bermoq, joy bermoq, zarba bermoq)

8. Koʻmakchi feʼl vazifasida affiksli ravishdosh shaklidagi yetakchi feʼl bilan birikadi -b/-ib=> 1) *harakatning birov, oʻzga uchun bajarilganligini ifodalaydi.* (Yangangga ayt: ilikdan osh

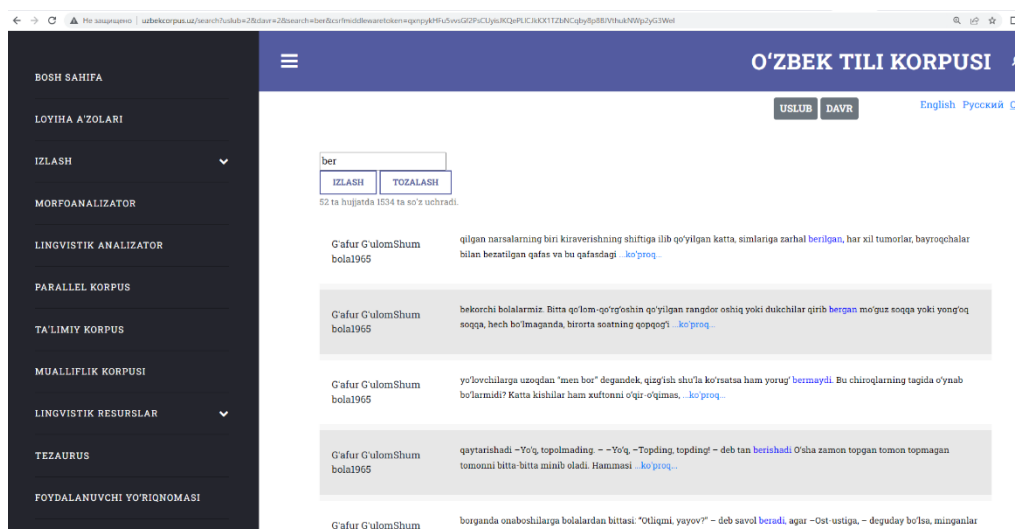
qilib bersin.Guldasta); 2) *yetakchi fe'ldan anglashilgan ma'noni kuchaytirib ifodalaydi*. U qo'lini siltab baqirib berdi.

11. Ko'makchi fe'l vazifasida affiksli ravishdosh shakllaridagi yetakchi fe'llar bilan birikib kelganda, ko'pincha -ver shaklida talaffuz qilinadi va quyidagi ma'nolarni ifodalaydi -a, -y=> 1) yetakchi fe'ldagi harakatning bajarilishi uchun qarshilik, monelik yo'qligini, u bemaolol bajarilishi mumkinligini anglatadi.

(Qatqlamasdan kosalarga suza ber. (A. Qodiriy, G'irvonlik Mallavoy); 2) yetakchi fe'ldagi harakatning davomlilik, takroriyiligini bildiradi.

12. Yetakchi fe'l sifatida qo'shma fe'llar tarkibiga kiradi, masalan: **berib tur, bera qol, berib yubordi**.

O'zbek tili elektron korpusida [3] ushbu so'zning bir nechta so'zlar bilan kelish holatini kuzatish mumkin va anglashiladiki, kontekstologik vaziyat so'zlarning lug'aviy maqomini belgilaydi.



Izohli lug'atda berilgan izohlar asosan, *bermoq* so'zidan oldingi so'z bilan semantik jihatdan bog'lanib, ushbu fe'lga sinonim bo'la olishi nuqtai nazaridan ma'no ifodalanishi qayd etilgan. Korpusga asoslangan leksikografik tadqiqotlarimiz shuni ko'rsatadiki, birikmalar tarkibida qo'llanishi o'zbek tili korpusi vositasida aniqlangan 1534 ta natijadan umumiy jihatdan quyidagi modellarni hosil qildi:

1. Ot+bermoq (V_{ber})=>
 - 1.1. yubormoq $\Leftrightarrow$ {xabar, signal+ V_{ber} }
 - 1.2. solmoq $\Leftrightarrow$ {quloq+ V_{ber} }
 - 1.3. qilmoq $\Leftrightarrow$ {yordam, maslahat, qaror, xalaqit, e'tibor, diqqat+ V_{ber} }
 - 1.4. chalmoq $\Leftrightarrow$ {xushtak+ V_{ber} }
 - 1.5. V+la $\Leftrightarrow$ {tormoz, nom, izoh, zahar, baho, nusxa+ V_{ber} }
 - 1.6. urmoq $\Leftrightarrow$ {qamchi+ V_{ber} }
 - 1.7. bag'ishlamoq $\Leftrightarrow$ {kuch, quvvat, umrini, taft, husn, ko'rk+ V_{ber} }
2. Sifat+bermoq(V_{ber})=> {yorug', chap, zo'r+ V_{ber} }
3. Fe'l₁+b/ib=> {qirib, suzib, o'qib...+ V_{ber} }
4. Fe'l₁+a/y=> {o'tira, ko'ra, eragasha...+ V_{ber} }
5. Fe'l₁+b/ib+fe'l₂+b/ib+bermoq=> olib kirib bermoq
6. Taqlid+bermoq=> {dov, hay...+ V_{ber} }

7. Ibora tarkibida=> {xiralik qila bermoq, yuragi dov bermoq, Xudo berardan qismasin, Xudo beraman desa, Xudo berib, sir boy bermoq, yuzini yerga bermoq, rasmiy tus bermoq, sovuq tus bermoq, jonini jabborga bermoq, bo'yni yor bermadi, taqdirga tan bermoq}.

Tahlil natijalar shuni anglatadiki, korpus asosida leksikografik tahlil so'zlarning semantik munosabatlar asosida muayyan maydonlar hosil qilishi natijasida matnni semantik tahlil qilishda muhim ahamiyat kasb etadi. Shuningdek, o'zbek tilini xorijiy yoki ikki til sifatida o'rganishda kollokativ birikmalar orqali ma'no farqlash imkoniyati yaratiladi.

Foydalanilgan adabiyotlar:

1. Abduraxmonova, N. Z. «Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref.» (2018).
2. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
3. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
4. 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.
5. 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;5(03):89-100.
6. A valency of English . A corpus based Anallysis of the Complementation Patterns of English Verb, Nouns and Adjectives Berlin, Walter de Gruyter, 2004
7. Lucien Tesniere Elements of Structural Syntax Amsterdam/Philadelphia 2015, - P. xlvii.

THE BASIC FEATURES OF THESAURUS DICTIONARIES

Saparova Mohira Fayzullayevna

Ma'mun university NEI, Faculty of Philology,
Department of Roman-German Philology, an English teacher

Annotation: This article indicates that Thesaurus is a valuable tool for expanding one's vocabulary and understanding the relationships between words and their meanings. It should be used thoughtfully and in conjunction with a good understanding of language and context. Besides, it obviously seen that the semantic and structural research of nouns and their specific peculiarities in this article.

Keywords: Thesaurus dictionaries, additional features, comprehensive tool, database, semantic research, structural research

Аннотация: В этой статье указывается, что тезаурус является ценным инструментом для расширения словарного запаса и понимания отношений между словами и их значениями. Его следует использовать вдумчиво и в сочетании с хорошим пониманием языка и контекста. Кроме того, очевидно, что семантическое и структурное исследование имен существительных и их специфических особенностей в данной статье.

Ключевые слова: тезаурусные словари, дополнительные возможности, комплексный инструмент, база данных, семантическое исследование, структурное исследование.

Annotatsiya: Ushbu maqolada tezaurus so'z boyligini kengaytirishi, shuningdek, so'zlar va ularning ma'nolari o'rtasidagi munosabatlarni tushunish uchun qimmatli vosita ekanligi ta'kidlangan. Ushbu lug'at keng mulohaza bilan, til va kontekstni yaxshi tushunish bilan birgalikda ishlatilishi kerak. Bundan tashqari, ushbu maqolada otlarning semantik va strukturaviy tadqiqi va ularning o'ziga xos xususiyatlari misollar yordamida keng yoritib beriladi.

Kalit so'zlar: tezaurus lug'atlar, qo'shimcha funksiyalar, murakkab vosita, ma'lumotlar bazasi, semantik tadqiqot, strukturaviy tadqiqot

A Thesaurus is a reference book or database that lists words in a language and groups them according to similarity in meaning, often with synonyms and antonyms. The purpose of a thesaurus is to help people find the right word to use in a particular context, to expand their vocabulary, and to improve their writing or speaking skills. Thesaurus can be organized in different ways, depending on their purpose or intended audience. Some thesaurus dictionaries are general and cover a broad range of topics, while others are specialized and focus on specific fields or subject areas, such as medicine, law, or technology. Some thesaurus are also designed for language learners, providing definitions, examples, and usage notes to help users better understand the meaning and use of words.

The word «thesaurus» comes from the Greek word «thesauros,» which means «treasure» or «storehouse.» In modern times, a thesaurus is seen as a treasure trove of words, offering a vast array of synonyms and related words that can help writers choose the right word for their intended meaning. [3, 45-48] One of the most widely used thesaurus in English is Roget's Thesaurus, its

words into categories based on their meaning and usage, and provides synonyms and related words for each category. For example, the category «joy» includes synonyms such as happiness, delight, and pleasure, while the category «fear» includes synonyms such as terror, dread, and panic. Other well-known English-language thesaurus include the Oxford Thesaurus of English, Longman Language Activator, Merriam-Webster Thesaurus, and Collins Thesaurus. Each of these thesaurus has its own unique features and benefits, and may be better suited to certain users or purposes. [4, 67]

Roget's Thesaurus, first published in 1852, is one of the most popular and widely used thesaurus in English. [5, 90-98] It groups words into categories based on their meaning and usage, and provides synonyms and related words for each category. Other well-known English-language thesaurus include the Oxford Thesaurus of English, Longman Language Activator, Merriam-Webster Thesaurus, and Collins Thesaurus. Thesaurus can also be found online, in the form of websites or software applications. Online thesaurus often provide additional features, such as search functions, audio pronunciations, and the ability to save and organize lists of a thesaurus is a reference book or database that lists words in a language and groups them according to similarity in meaning. It is an essential tool for writers, speakers, and language learners who want to expand their vocabulary and improve their communication skills.

Thesaurus can also be found online, in the form of websites or software applications. Online thesaurus dictionaries often provide additional features, such as search functions, audio pronunciations, and the ability to save and organize lists of words. Some online thesaurus are also integrated with other language resources, such as dictionaries and grammar guides, making them a comprehensive tool for language learners. In addition to providing synonyms and related words, thesaurus can also help users understand the nuances of meaning and usage between different words.

For example, two words may be synonyms, but one may be more formal or technical than the other, or may have connotations that the other does not. By consulting a thesaurus, users can gain a deeper understanding of the subtleties of language and choose the right word for their intended meaning. In general, a thesaurus is an essential tool for anyone who wants to improve their writing, speaking, or language learning skills. With its vast array of synonyms and related words, a thesaurus can help users expand their vocabulary and choose the right word for their intended meaning. Whether in print or online, a thesaurus is a treasure trove of words that can enhance communication and enrich language learning. words. In a thesaurus, nouns are organized by their meanings and grouped together with other nouns that have similar meanings, called synonyms. A thesaurus is a reference book or database that lists groups of words that are related in meaning. When you look up a word in a thesaurus, you will find a list of synonyms for that word, as well as related words and phrases. For example, if you look up the word «happy» in a thesaurus, you might find synonyms such as «joyful», «content», and «pleased». You might also find related words like «smile», «laugh», and «celebrate». These related words are not synonyms, but they are related in meaning and can help you to better understand the nuances of the word you are looking up. Thesauruses can be useful tools for writers, students, and anyone looking to expand their vocabulary. They can help you to find more precise or interesting words to use in your writing, and can also help you to understand the subtle differences in meaning between similar words.

The semantic structure of a noun refers to the underlying meaning or conceptual structure of the noun. Nouns can be analyzed in terms of their semantic features, which are the basic components of meaning that make up a word's definition. These features can include things like shape, size, color, texture, function, and location. For example, the noun «apple» can be analyzed in terms of its semantic features. Some of its features might include round shape, red or green color, sweet taste, and the ability to be eaten. These features help to define what an apple is and distinguish it from other fruits.

Nouns can also be analyzed in terms of their semantic roles, which are the roles that participants in a sentence play in relation to the noun. These roles include things like agent, patient, theme, and location. For example, in the sentence «The cat chased the mouse», «cat» is the agent (the one doing the action), «mouse» is the patient (the one being acted upon), and «chased» is the action verb. Understanding the semantic structure of nouns can be useful for language learners and researchers, as it can help them to better understand the meaning and use of words in a language. It can also be helpful for natural language processing (NLP) tasks such as text classification, sentiment analysis, and information retrieval. [7, 98-99] In a thesaurus, nouns are typically organized by their meanings and grouped together with other nouns that have similar meanings, called synonyms. The semantic structure of the nouns in a thesaurus is not explicitly shown, but it can be inferred based on the grouping of related words. For example, if you look up the word «computer» in a thesaurus, you might find synonyms such as «PC», «desktop», and «laptop». These synonyms all refer to similar types of computing devices, and are grouped together because they share a common semantic structure. Other related words might include «software», «hardware», «keyboard», and «mouse», which are not synonyms but are related in meaning and can help to provide a more complete understanding of the semantic structure of the word «computer». Overall, the semantic structure of nouns in a thesaurus is not explicitly stated, but can be inferred based on the organization of related words and their meanings. The thesaurus can be a useful tool for understanding the semantic relationships between nouns and expanding one's vocabulary.

Hierarchical words in a thesaurus refer to words that are arranged in a hierarchical structure based on their meanings. This means that words are grouped together based on broader categories and more specific subcategories. The hierarchy of words in a thesaurus can help users to understand the relationships between words and their meanings. For example, a thesaurus might have a hierarchy of words related to the concept of «food». At the top of the hierarchy might be the broad category «food», which would include subcategories such as «fruit», «vegetables», «meat», and «dairy». Each of these subcategories would then have their own subcategories, such as «apples», «bananas», and «oranges» under the category of «fruit». The hierarchical structure of words in a thesaurus can be useful for language learners and researchers, as it can help them to understand the relationships between words and their meanings. It can also be helpful for natural language processing (NLP) tasks such as text classification, information retrieval, and concept mapping.

Thus, the hierarchical structure of words in a thesaurus can help users to understand the relationships between words and their meanings, making it a useful tool for expanding one's vocabulary and understanding the nuances of language. Structural research of nouns involves analyzing the internal structure of a noun, as well as its relationships to other words and concepts. One approach to studying the structure of nouns is through the use of morphological analysis.

Morphology is the study of the structure of words and how they are formed from smaller units called morphemes. For example, the noun «unhappiness» can be broken down into three morphemes: «un-», «happy», and «-ness». By analyzing the morphological structure of a noun, researchers can gain insights into its meaning and relationships to other words.

Another approach to studying the structure of nouns is through the use of semantic analysis. Semantic analysis involves analyzing the meaning of words and how they relate to other words and concepts. For example, the noun «apple» can be analyzed in terms of its semantic features, such as its shape, color, taste, and function. By analyzing the semantic structure of a noun, researchers can gain insights into how it is used in language and how it relates to other words and concepts.

Overall, the structural research of nouns involves analyzing their internal structure and relationships to other words and concepts. This can be done through various approaches, such as morphological analysis and semantic analysis, and can provide valuable insights into the meaning and use of words in language. Thesaurus dictionaries are very helpful for finding out the meaning and semantic- structural research of a word, in my view, we can expand our vocabulary knowledge with the help of the thesaurus dictionaries.

References:

1. Abdurakhmonova, N. Z. «Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aforef.» (2018).
2. 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.
3. Abdurakhmonova N.Z, 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.
4. Alessandro Agostini, Timur Usmanov, Ulugbek Khamdamov, Nilufar Abdurakhmonova, and Mukhammadsaid Mamasaidov. 2021. UZWORDNET: A Lexical-Semantic Database for the Uzbek Language (<https://aclanthology.org/2021.gwc-1.2>). In Proceedings of the 11th Global Wordnet Conference, pages 8–19, University of South Africa (UNISA). Global Wordnet Association.
5. ACL Anthology (<https://aclanthology.org/2021.gwc-1.2/>)
6. George Davidson, ed., Thesaurus of English words and phrases (150th Anniversary Edition), Penguin, 2002, ISBN 0141004428, p. 454
7. George Crabb, English Synonyms Explained, in Alphabetical Order with Copious Illustrations and Examples Drawn from the Best Writers, 2nd edition, London 1818 full text
8. Peter Mark Roget, Thesaurus of English words and phrases, classified so as to facilitate the expression of ideas, 1853, V.ii.iv, p. 204
9. Roget, P. M. (1982). «Preface to the first edition, 1852». In Lloyd, Susan M. (ed.). Roget's Thesaurus of English Words and Phrases (New ed.). Harlow: Longman. p. xix.

VERBS IN THESAURUS

Xudoyberganov Sirojiddin Meyliboyevich

Ma'mun University NEI, Faculty of "Philology"

Department of "Roman-German Philology", senior Teacher

Annotation: This article is dedicated to the problem of searching a definite part of speech - verb and its categories. We tried to analyze the forms of different subdivisions and their usage with some examples. Thesaurus is famous for searching various synonyms and related words of verbs.

Keywords: Thesaurus, verb, synonym, linguistic analyses, auxiliaries, modals, complementation.

Annotatsiya: Ushbu maqola nutqning aniq qismini – fe'l va uning kategoriyalarini izlashda yuzaga keladigan muammolarga bag'ishlangan. Bunda biz turli bo'linmalarining shakllari va ulardan foydalanishni ba'zi misollar bilan tahlil qilishga harakat qilganmiz. Tezausur turli xil sinonimlar va fe'llarning tegishli so'zlarini qidirish bilan mashhurdir.

Kalit so'zlar: Tezausur, fe'l, sinonim, lingvistik tahlil, ko'makchi, modal, to'ldiruvchi.

Аннотация: Данная статья посвящена проблеме поиска определенной части речи - глагола и его категорий. Мы попытались проанализировать формы различных подразделений и их использование на некоторых примерах. Тезаурус известен поиском различных синонимов и родственных слов глаголов.

Ключевые слова: тезаурус, глагол, синоним, лингвистический анализ, вспомогательные формы, модальные формы, дополнение.

Thesaurus is a reference tool that provides synonyms and antonyms for words. While Thesaurus can be useful for finding alternative words that have similar or opposite meanings to a given verb, it does not provide a structural analysis of the verb or its usage in context. [6, 38-40]

For example, if you were to look up the verb «run» in Thesaurus, you might find synonyms such as jog, sprint, dash, and race, as well as related words like exercise, workout, and fitness. However, a thesaurus would not provide information about the tense, aspect, mood, or complementation patterns of the verb, or how it is used in different syntactic or discourse contexts.

Therefore, while Thesaurus can be a helpful resource for expanding your vocabulary and finding alternative words, it should not be relied upon solely for understanding the structural and functional properties of a verb in English. Other resources, such as grammar books, linguistic analyses, and language corpora, can provide more detailed and nuanced information about the properties and usage of verbs in the English language.

The structure of a verb in English refers to its form, tense, aspect, mood, voice, and complementation patterns, which are essential for understanding its meaning and usage in different contexts. [1, 56-59] These properties can be analyzed and described using linguistic frameworks such as syntax, semantics, and pragmatics which is not discussed in thesaurus. Besides, thesaurus typically does not provide information about modals in English. Modals are a type of auxiliary

verb that express a speaker's attitude or degree of certainty towards a particular situation, and they have unique functions and meanings that are not easily interchangeable with other words.

For example, the modal verb «can» expresses ability or possibility, while «should» expresses obligation or recommendation, and «might» expresses uncertainty or possibility. While Thesaurus might provide synonyms for the main verb in a sentence, it would not provide alternatives for the modal verb, as these words have specific meanings and functions that are not easily interchangeable with other words.

Together with Thesaurus typically provides similar and opposite meanings for words, including gerunds. However, it may not provide specific information about the grammatical properties of gerunds, such as their function as verbal nouns or their ability to take direct objects. It is important to note that Thesaurus is a reference tool for finding alternative words with similar meanings, but it should not be relied upon as a comprehensive grammar guide.

When searching for verb synonyms or antonyms in Thesaurus, it is important to keep in mind that not all synonyms or antonyms are interchangeable in every context. On the one hand different verbs may have different shades of meaning or connotations, and may be used in different ways depending on the context on the other hand antonyms are words that have opposite meanings to a given word, and can be useful for adding variety and contrast to writing or speech.

To use Thesaurus effectively when searching for verb synonyms, it is helpful to have a clear understanding of the context in which the verb will be used. This can help to identify the most appropriate synonyms, and to avoid using words that have different connotations or meanings that are not appropriate for the context. On the contrary, while looking for antonyms of verbs in Thesaurus, it is important to keep in mind that not all verbs have antonyms, and that some verbs may have multiple antonyms depending on the context. For example, the verb «to love» could have antonyms such as «to hate» or «to despise», depending on the context in which the word is used.

When consulting Thesaurus for verb synonyms or antonyms, it is also important to consider the grammatical properties of the verbs, including their tense, aspect, voice, and mood. Some verbs may be more appropriate for certain grammatical contexts than others or some verbs may have antonyms that are grammatically incorrect or inappropriate for the intended meaning. [2, 9-10]

Overall, Thesaurus can be a useful tool for finding alternative verbs with similar or opposite meanings to a certain verb but it should be used alongside other resources, such as a dictionary or grammar guide, to ensure that the chosen words are appropriate for the intended context and grammatically correct.

References:

1. Abdurakhmonova N. Modeling Analytic Forms of Verb in Uzbek as Stage of Morphological Analysis in Machine Translation. Journal of Social Sciences and Humanities Research. 2017;5(03):89-100.
2. *American Heritage Dictionary of the English Language*, 5th edition, Houghton Mifflin Harcourt 2011, p. xxvii
3. Edwin L. Battistella, «Beware the thesaurus», OUP blog, «Oxford University Press's Academic Insights for the Thinking World», February 11, 2018

4. N. Z. Abdurakhmonova, A. S. Ismailov and D. Mengliev, «Developing NLP Tool for Linguistic Analysis of Turkic Languages,» 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation, 2022, pp. 1790-1793, doi: 10.1109/SIBIRCON56155.2022.10017049.
5. 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.
6. 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.
7. Roget, Peter. 1852. *Thesaurus of English Language Words and Phrases*.
8. *The Merriam-Webster thesaurus*. Springfield, Massachusetts: Merriam-Webster. 2005.
9. Werner Hüllen, «Roget's Thesaurus, deconstructed» in *Historical Dictionaries and Historical Dictionary Research*, papers from the *International Conference on Historical Lexicography and Lexicology*, University of Leicester, 2002, Max Niemeyer Verlag 2004, p. 83–94

KOMPYUTER LEKSIKOGRAFIYASIDA ASSOTSIATIV MAYDON
BIRLIKLARINING GENDEROLOGIK TADQIQI

Jo‘raqulova Madina Subhonovna,
O‘zbekiston Milliy universiteti tadqiqotchisi

Xoliyorova Gulruxsor
O‘zbekiston Milliy universiteti tadqiqotchisi

Annotatsiya: Ushbu maqolada assotsiativ maydon birliklarining gender tadqiqi haqida fikr yuritilgan bo‘lib, “oila” konseptini voqelantiruvchi leksik birliklar gender nuqtayi nazaridan tahlil qilingan. Ushbu masalarni o‘rganish asosida o‘zbek tilida “oila” so‘zi bilan assotsiativ bog‘lanuvchi birliklarga aniqlik kiritilgan. “Oila” arxisemali maydon birliklari gender tahlilga tortilgan. Ayollar va erkaklarda “oila” leksemasi bo‘yicha hosil qilingan assotsiativ maydon birliklari tahlil qilingan va gender farqlari ko‘rsatib berilgan. Bu o‘rinda “Oila” leksemasi bilan assotsiativ bog‘lanuvchi birliklarning hosil bo‘lishida lingvistik va ekstralingvistik, grammatik-morfologik omillarning ta’siri izohlab berilgan.

Kalit so‘zlar: assotsiativ birliklar, assotsiativ aloqa, assotsiativ maydon, leksik assotsiatsiya, individual assotsiatsiya, genderlekt.

Annotation: This article is discussed the gender research of associative field units and lexical units embodying the concept of “family” are analyzed from a gender point of view. Based on the study of these issues, the units associated with the word “family” in the Uzbek language have been clarified. The archetype “family” field units are subjected to gender analysis. The associative field units formed by the lexeme “family” in women and men were analyzed and gender differences were shown. Here, the influence of linguistic and extralinguistic, grammatical and morphological factors in the formation of associative units with the lexeme “family” is explained.

Keywords: associative units, associative relation, associative field, lexical association, individual association, genderlect.

Аннотация: В данной статье рассматриваются гендерные исследования единиц ассоциативного поля и лексических единиц, воплощающих концепт “семья”, анализируются с гендерной точки зрения. На основе изучения этих вопросов уточнены единицы, связанные со словом “семья” в узбекском языке. Гендерному анализу подвергаются архетипические “семейные” полевые единицы. Проанализированы единицы ассоциативного поля, образованные лексемой “семья” у женщин и мужчин, показаны гендерные различия. Здесь объясняется влияние языковых и экстралингвистических, грамматических и морфологических факторов на образование ассоциативных единиц с лексемой “семья”.

Ключевые слова: ассоциативные единицы, ассоциативная связь, ассоциативное поле, лексическая ассоциация, индивидуальная ассоциация, гендерлект.

Gender lingvistikasida assotsiativ tajribalarning qo'llanilishi ayollar va erkaklarda vujudga keladigan verbal assotsiatsiyalarning bir-biridan farqi va ularning qanday omilga ko'ra yuzaga chiqishini o'rganishda qo'l keladi. Sababi, har qaysi shaxsning hayoti mobaynida to'plagan tajribasi va reprezentativ tizim modalliklari assotsiatlarning turlicha bo'lishiga sababchi omil sifatida e'tirof etish mumkin. [Залевская А. 1999:75]. Har qaysi kishida har bir so'zga taalluqli o'ziga xos individual tajriba shakllanadi. Kishi ongida so'zlar mazmuniy guruhlarini tashkil etadi. Ushbu assotsiativ tajriba ma'lum shaxs ongining psixologik mazmunini obyektiv tarzda o'rganishga yordam beradi.

N.B. Krushevskiy assotsiativ munosabatni quyidagicha izohlaydi: "Har qanday so'z o'xshashlik belgisiga ko'ra boshqa so'zlar bilan bog'liq bo'ladi, bu o'xshashlik nafaqat tashqi, tovush tomondan, morfologik jihatdan, balki ichki, semasiologik jihatga ko'ra ham ongimizda o'ziga o'xshash boshqa so'zlarni qo'zg'atishga va ular bilan birga qo'zg'alishga qodir bo'ladi" [Lutfullayeva.D. 2017:86]. Biror bir so'z (stimul so'z)ni eshitganimizda yoki ko'rganimizda ma'lum birliklar beixtiyor ravishda bir-birini eslata boshlaydi. Mana shu "eslatish" jarayoni assotsiativ birliklarning paydo bo'lish asosi hisoblanadi.

1.1. Ayol sinaluvchilarda hosil qilingan assotsiativ maydon: tarkibi va gender xususiyatlari.

Biz Oila so'zining ayollar ongida qanday assotsiativ birliklar bilan o'zaro maydon hosil qila olishini kichik tajribada sinab ko'rdik va quyidagi birliklar paradigmasiga egaligini aniqladik:

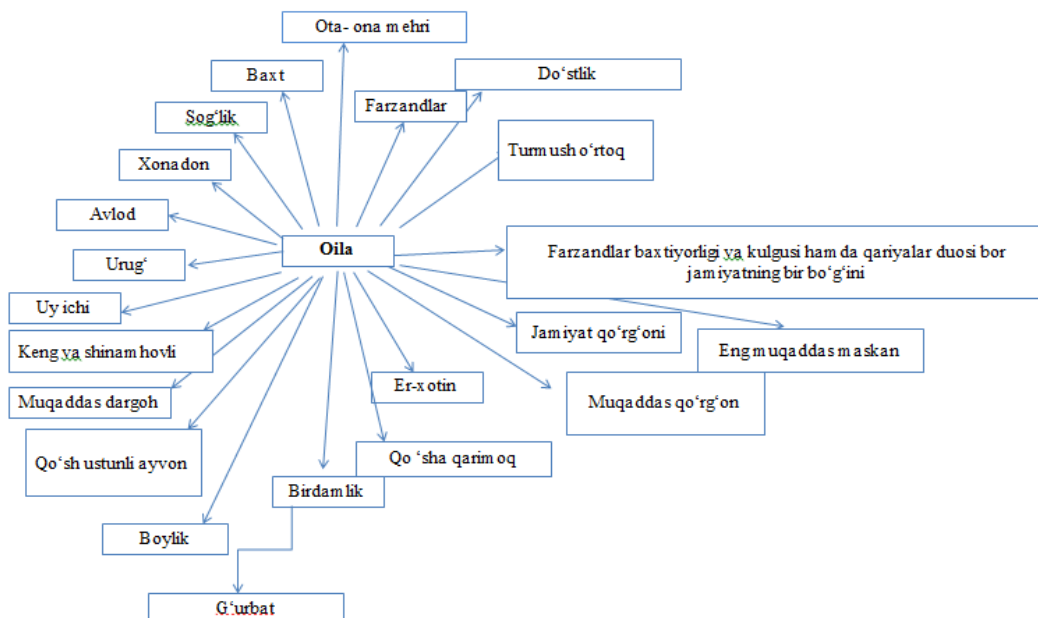
baxt; sog'lik; boylik; muqaddas dargoh; farzandlar(3); turmush o'rtoq; eng muqaddas maskan; ota-ona mehri; farzandlar baxtiyorligi va kulgusi hamda qariyalar duosi bor jamiyatning bir bo'g'ini; qo'sha qarimoq, g'urbat, muqaddas go'sha; Jamiyat qo'rg'oni, farzand; keng va shinam hovli; birdamlik; do'stlik; uy ichi; qo'sh ustunli ayvon; xonadon; avlod; urug'; er-xotin.

Bu tajribada oila stimul so'zi biror o'rinda takror uchramadi; so'z shaklidagi assotsiatsiyalar 11 ta (*baxt, sog'lik, boylik, farzandlar, farzand, birdamlik, do'stlik, g'urbat, xonadon, avlod, urug'*); qo'shma so'z shaklidagi assotsiatsiyalar 2 ta (*turmush o'rtoq, qo'sha qarimoq*); juft so'z shaklidagi assotsiatsiyalar 1 ta (*er-xotin*); so'z birikmasi shaklidagi assotsiatsiyalar 8 ta (*muqaddas dargoh, eng muqaddas maskan, ota-ona mehri; muqaddas go'sha, Jamiyat qo'rg'oni, keng va shinam hovli, uy ichi, qo'sh ustunli ayvon*); gap shaklidagi assotsiatsiyalar 1 ta (*farzandlar baxtiyorligi va kulgusi hamda qariyalar duosi bor jamiyatning bir bo'g'ini*)ni tashkil qildi. Mazkur birliklarning oila stimul so'zi bilan munosabatlari turlicha. Xususan, *baxt, sog'lik, boylik,, birdamlik, do'stlik* assotsiatsiyalari insonning voqelikka munosabati asosida yuzaga chiqqan. *Muqaddas dargoh, eng muqaddas maskan, farzandlar baxtiyorligi va kulgusi hamda qariyalar duosi bor jamiyatning bir bo'g'ini, muqaddas go'sha, Jamiyat qo'rg'oni, qo'sh ustunli ayvon, uy ichi* leksemalari esa oila stimul so'zi bilan predmetning makondagi aloqadorligi o'xshatish munosabati asosida hosil bo'lgan. *Ota-ona mehri* o'xshatish munosabatini aks ettirgan bo'lsa, *qo'sha qarimoq* so'zi esa oila stimul so'zi bilan sabab-natija munosabatini hosil qilgan. *Farzandlar, turmush o'rtoq, er-xotin* assotsiatsiyalari o'zaro stimul so'z bilan butun-qism munosabatini yuzaga chiqargan. *Xonadon, avlod, urug'* assotsiatsiyalari esa stimul so'z bilan paradigmatic munosabatni yuzaga chiqargan. *G'urbat* so'zi esa insonning voqelikka individual munosabatini ifodalab kelgan.

O'tkazilgan tajribadagi assotsiativ maydon birliklari munosabatini chizmada quyidagicha tasvirlash mumkin:

1-sxema. “Oila” stimol soʻziga ayol sinaluvchilarda hosil qilingan assotsiativ maydon, tarkibi.

1.2. Erkak sinaluvchilarda hosil qilingan assotsiativ maydon: tarkibi va gender xususiyatlari.

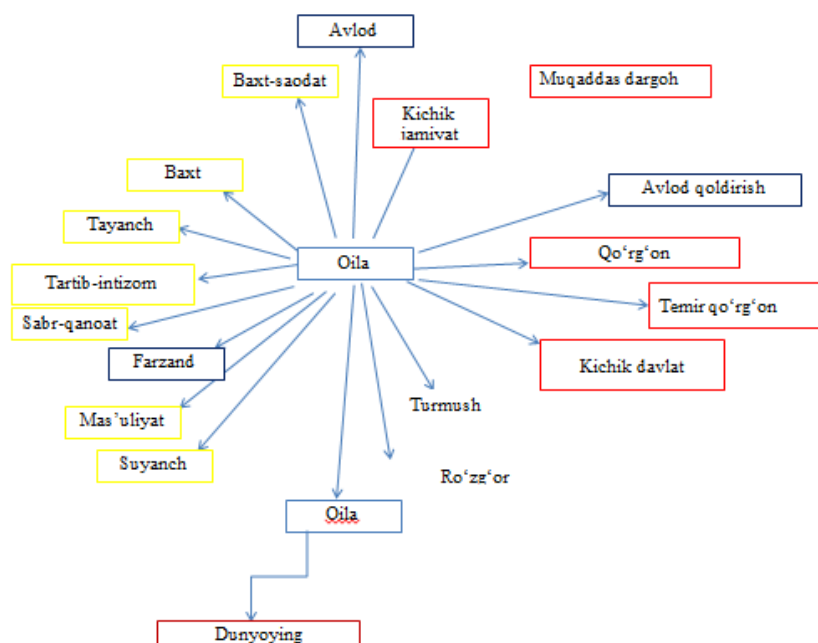


Biz *Oila* stimol soʻzining erkaklar ongida qanday assotsiativ birliklar bilan oʻzaro maydon hosil qila olishini kichik tajribada sinab koʻrdik va quyidagi birliklar paradigmasiga egaligini aniqladik:

avlod qoldirish, muqaddas dargoh, kichik jamiyat, kichik davlat; baxt-saodat, temir qoʻrgʻon, baxt(2), qoʻrgʻon, dunyoying, farzand, masʼuliyat, roʻzgʻor, turmush, tartib-intizom, tayanch, avlod, sabr-qanoat.

Bu tajribada *oila* stimol soʻzi biror oʻrinda takror uchramadi; soʻz shaklidagi assotsiatsiyalar 9 ta (*baxt, qoʻrgʻon, dunyoying, farzand, masʼuliyat, roʻzgʻor, turmush, tayanch, avlod*); qoʻshma soʻz shaklidagi assotsiatsiyalar 1 ta (*avlod qoldirmoq*); juft soʻz shaklidagi assotsiatsiyalar 3 ta (*baxt-saodat, tartib-intizom, sabr-qanoat*); soʻz birikmasi shaklidagi assotsiatsiyalar 4 ta (*muqaddas dargoh, kichik jamiyat, kichik davlat, temir qoʻrgʻon*); gap shaklidagi assotsiatsiyalar 0 ni tashkil qiladi.

Mazkur birliklarning *oila* stimol soʻzi bilan munosabatlari turlicha. Xususan, *baxt, baxt-saodat, dunyoying, masʼuliyat, tartib-intizom, sabr-qanoat, tayanch* assotsiatsiyalari insonning voqelikka munosabati asosida yuzaga chiqqan. *Muqaddas dargoh, temir qoʻrgʻon, qoʻrgʻon soʻz birikmalari* esa *oila* stimol soʻzi bilan predmetning makondagi aloqadorligi oʻxshatish munosabati asosida hosil boʻlgan. *Avlod qoldirish* soʻzi esa *oila* stimol soʻzi bilan sabab-natija munosabatini hosil qilgan. *Kichik jamiyat, kichik davlat, avlod* assotsiatsiyalari esa stimol soʻz bilan paradigmatic munosabatni yuzaga chiqargan. *roʻzgʻor, turmush* soʻzi esa predmetning oʻxshashlik munosabatini ifodalab kelgan. Jamoaviy assotsiativ maydon tekshiriluvchilar tomonidan qayd etilgan soʻzlarga asosan shakllandi. Bunda koʻp qayd etilgan soʻzlar maydon yadrosini tashkil qilgan boʻlsa, yagona reaksiyalar maydon periferiyasiga joylashtirildi. Oʻtkazilgan tajribadagi assotsiativ maydon birliklari munosabatini chizmada quyidagicha tasvirlash mumkin:



2-sxema. “Oila” stimol soʻziga erkak sinaluvchilarda hosil qilingan assotsiativ maydon, tarkibi.

Assotsiativ maydon stimol soʻz vazifasidagi leksik birlikning maʼno xususiyatlarini oʻrganishda muhim oʻrin tutadi. Assotsiativ maydon stimol soʻzning lugʻaviy maʼnosi, lugʻatlarda aks etmagan ayrim konnotativ, uslubiy, pragmatik, dialektal maʼnolari, nutqiy qoʻllanishdagi xususiyatlari, koʻchma maʼnolarini aniqlashga imkon beradi. Assotsiativ tajriba materiallari kompyuter leksikografiyasi uchun manba sifatida xizmat qilishi mumkin. [Abdurakhmonova N. 2022:1791]. Assotsiativ maydon uch zamonning sinxron harakatini yuzaga chiqaradi. Ishda assotsiativ maydon birliklarining gender farqlanishi “oila” arxisemasiga taalluqli ayollar va erkaklarga xos jamoaviy assotsiativ maydonlar orqali tahlil qilindi. Ayollar va erkaklar ongida biz *oila* soʻzining qanday assotsiatsiyalar hosil qilishini koʻrib oʻtdik. Yuqoridagi tahlillar asosida har ikkala assotsiativ maydonning umumiy aʼzolari sifatida quyidagi birliklarni keltirish mumkin: *baxt-saodat*, *muqaddas dargoh*, *farzand*. Demak, bu tushunchalar *oila* stimol soʻzi bilan assotsiativ munosabatga kirisha oladigan umumiy tushunchalar. Darhaqiqat, oʻzbek xalqi uchun *oila* bu-muqaddas dargoh, *oila* bu-baxt, oilaning ustuni farzanddir. Erkaklarda hosil qilingan assotsiativ maydonda *oila* stimulining javob reaksiyalari oʻz maʼnosida qoʻllangani, ayollarda esa assotsiativ birliklarning koʻproq ramziylik kasb etganining guvohi boʻlamiz (buning bosh sababi – ayollar va erkaklar tasavvurlarining oʻziga xosligidadir). Har ikkala maydonda assotsiatsiyalarning hosil boʻlish omillari oʻxshash boʻlsa ham, mazmunan turlicha yondashilgan. Ikkala maydonda ham respondentlarning oʻziga xos xarakter-xususiyatlari akslanib turibdi. Soʻzlarning assotsiatsiyalari nafaqat vaziyatga aloqador, balki effekt, ichki his-hayajon, tuygʻu jarayoni ham boʻlishi mumkin. [Мартинович Г. 1993:96]. Shuning uchun assotsiativ maydon tuzilishida turli individual assotsiatsiyalarni ham uchratamiz. Oʻtkazilgan tajribada ayollarning javob reaksiyalarida “*qoʻsh ustunli ayvon, qoʻsha qarimoq yoki gʻurbat*” kabi birikuvlar nutq egasining voqelikka individual munosabatini alohida namoyon qilgan boʻlsa, erkaklarda esa “*temir qoʻrgʻon, tartib-intizom, butun dunyo*” birliklari bilan hosil qilingan assotsiatsiyalar bevosita shaxsiy individuallikni yuzaga chiqargan. Yaʼni har bir shaxsda stimol soʻzga nisbatan umumiy assotsiatsiyalardan tashqari shaxsiy, individual munosabat ham mavjuddir. Ayol

sinaluvchilarda hosil qilingan assotsiativ maydon birliklari soʻz birikmasi va gap shaklidagi assotsiatsiyalarning koʻp qoʻllanganligi bilan ajralib turadi. Erkaklarda hosil qilingan assotsiatsiyalarning aksariyati soʻz va soʻz birikmasi shaklidagi assotsiatsiyalardir.

Ayollarda hosil qilingan assotsiatsiyalarda voqelikka nisbatan salbiy (gʻurbat) va ijobiy (eng muqaddas maskan) baho munosabatlarining yuzaga chiqqanligini koʻrishimiz mumkin. Erkaklarda esa leksemalarning semantikasiga asoslangan munosabatlari (turmush, roʻzgʻor kabi) yuzaga chiqqan.

Individual assotsiatsiyaning genderologik xususiyatlari mavjud boʻlib, bular ijtimoiy, milliy, biologik-fiziologik koʻrsatkichlar milliy mentalitetga, ular yashayotgan davrga, jamiyatga chambarchas bogʻliqdir. Tajriba natijalari shuni koʻrsatdiki, erkaklar nutqi ayollarnikidan qisqa tuzilishga ega, aksincha, ayol nutqi erkaknikiga nisbatan uzundan-uzoq shaklda namoyon boʻladi; ayolning soʻz qoʻllash bilan bogʻliq miqdoriy koʻrsatkichi erkaklarnikidan yuqori darajani koʻrsatdi; ayollarning nutqida erkaklarnikiga qaraganda salbiylikka yon bosish koʻproq uchrashining guvohi boʻldik.

Assotsiatsiyalarning hosil boʻlishida ayollar va erkaklarning jamiyatda, oilada tutgan oʻrni, vazifasi kabi omillar ham taʼsir koʻrsatgan. Masalan: ayollarda “oila” stimuli soʻziga nisbatan oila aʼzolari, oilaning muqaddas ekanligi kabi tushunchalar anglashilsa, erkaklarda turmush, roʻzgʻor kabi assotsiatsiyalar, oila oldidagi masʼuliyat, oilaga tayanch, suyanch boʻlish kabi tushunchalar gavdalangan. Yuqoridagi fikrlarimizga qoʻshimcha qilib shuni aytish joizki, oʻzbek eriga oilaparvarlik, halollik, botirlik, mardlik, toʻgʻrisoʻzlik kabi xislatlar xos boʻlsa; oʻzbek ayoliga esa mehribonlik, xushmuomalalik, donolik, or-nomus, farosat, sabr-qanoatlik, mehmondoʻstlik, mehnatsevarlik, bolajonlik kabi fazilatlar xos.

Demak, assotsiativ tajriba faqat stimuli soʻzning turfa maʼnolarini ochib beribgina qolmay, inson shaxsiyatini ham aniqlab beradi. Shu kabi tahlillar asosida insonlarning voqelikka munosabati, uning tafakkuri koʻlami haqida maʼlumotga ega boʻla olamiz. Demak, nutqning shakllanishida shaxsning assotsiativ tafakkuri mahsullari boʻlmish verbal assotsiatsiyalar tayanch birliklar sifatida ishtirok etadi. Assotsiativ tajribada berilgan assotsiativ birliklar tahlili insonlarning voqelik haqidagi tasavvurlari, uni qay tarzda idrok etishi haqida ham xulosa chiqarishga imkon beradi.

Foydalanilgan adabiyotlar:

1. Abdurakhmonova N.Z., Ismailov A. and Mengliyev D., “Developing NLP Tool for Linguistic Analysis of Turkic Languages”, 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation, 2022, pp. 1790-1793, doi: 10.1109/SIBIRCON56155.2022.10017049.
2. Abdurakhmonova N., Alisher I. 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. Залевская А.А. Введение в психолингвистику: Учебник для студентов высших учебных заведений. – М.: Российский гос. гуманитар. ун-т, 1999.
4. Лутфуллаева Д. Ассоциатив тилшунослик назарияси. – Т.: Меринос, 2017. –140 б.
5. Мартинович Г.А. Опыт комплексного исследования данных ассоциативного эксперимента // Вопросы психологии. - М., - 1993. - № 2. - С. 93-99.

TILSHUNOSLIK AXBOROT TIZIMIDA TEZAURUSDAN FOYDALANISH
ASOSLARI

Babajanova Lobar Jamoladdin qizi

Urganch davlat universiteti

Tarjima nazariyasi va amaliyoti kafedrası

assistent o'qituvchisi

Annotatsiya: Ushbu maqolada ilm-fan taraqqiyotida tezaurusdan foydalanishning ilmiy-nazariy asoslari va shakllanish bosqichlari haqida fikr-mulohazalar yuritilgan. Shuningdek, tezaurusning turlari, lug'atchilikka kompyuter texnologiyalarini tatbiq etish va tezaurus tushunchasini boshqa lug'at turlaridan farqlab olishga alohida e'tibor qaratilgan.

Kalit so'zlar: assotsiativ, iyerarxik, tezaurus, lingvistika, SNOMED, WORDNET, ISO, VINITI, MSC.

Annotation: this article discusses the scientific-theoretical basis for the use of thesaurus in the development of Science and the stages of formation. Special attention is also paid to the types of thesaurus, the application of computer technology to lexicography and the differentiation of the concept of thesaurus from other types of dictionaries.

Keywords: associative, hierarchical, thesaurus, linguistics, SNOMED, WORDNET, ISO, VINITI, MSC.

Аннотация: в данной статье рассмотрены научно-теоретические основы и этапы становления использования тезауруса в развитии науки. Также особое внимание уделяется видам тезауруса, внедрению компьютерных технологий в словарь и отличию понятия тезаурус от других видов словаря.

Ключевые слова: ассоциативный, иерархический, тезаурус, Лингвистика, SNOMED, WORDNET, ISO, VINITI, MSC.

Kirish.

Umuman olganda, axborot texnologiyalarining, shuningdek, axborotni qayta ishlashni uzatish sohasidagi texnologiyalarning rivojlanishi ilmiy va o'quv jarayonining deyarli barcha bosqichlarini tashkil etish uchun tubdan yangi imkoniyatlarning paydo bo'lishiga olib keldi, bu esa o'z navbatida o'qituvchilarning axborot ehtiyojlari sifatli o'sishiga olib keldi. Kompyuter bilan qurollangan, har kuni Internet imkoniyatlaridan foydalanadigan zamonaviy talaba o'quv jarayonining an'anaviy rejimi va o'quv materiallarining odatiy formatlari bilan qoniqmaydi, masalan: *darsliklar, kitoblar yoki tekis matnli fayllar*. Bugungi kunda o'quv materiallari o'quvchiga ko'plab raqamli formatlarda taqdim etilishi mumkin, ular turli xil qidiruv va tasniflash xizmatlari tomonidan qo'llab-quvvatlanishi kerak. Mavjud axborot resurslarini foydalanuvchi ehtiyojlariga muvofiq tizimlashtirish va tasniflash ham ilmiy, ham ta'lim faoliyatini qo'llab-quvvatlashning eng muhim vazifalaridan biridir. Ilmiy va, ayniqsa, o'quv faoliyati jarayonida adabiy manbalar, turli xil materiallar va hujjatlar bilan ishlash ko'p vaqt va kuch sarflaydi, kerakli hujjatlarni qidirish, hujjatlarni vazifaga muvofiq tizimlashtirish va tasniflash. Zamonaviy foydalanuvchilarning axborot ehtiyojlarini qondirish uchun axborotni qidirish va tasniflashning murakkab funksiyalarini qo'llab-quvvatlash, shuningdek, toifalar va tasniflagichlarning lug'atlari

bo'yicha manbalarni ko'rish kerak. Eng muhim vazifa resurslarni tizimlashtirish vazifasidir, uni hal qilish uchun foydalanuvchini qiziqtirgan, tanlangan tor mavzuni qamrab oladigan mantiqiy-semantik toifalar va asosiy atamalar tarkibini aniqlash kerak. Odatda, mavzu yo'nalishi o'rganilayotgan o'quv kursi yoki ma'lum bir kurs mavzusi bilan cheklanadi. Hozirgi vaqtda foydalanuvchilarning axborot ehtiyojlarini u yoki bu darajada qondiradigan yetarlicha kuchli axborot tizimlari mavjud. Shu bilan birga, aksariyat tizimlarning asosiy kamchiliklari resurslar bilan tahliliy ishlarni amalga oshirish va har bir tizim ichida ham, tashqi tizimlar bilan ham resurslarning integratsiyasini ta'minlash imkoniyatlarining cheklanganligi, ko'pincha, xalqaro standartlar va tavsiyalar hisobga olinmaydi. Bu ilmiy va ta'lim faoliyati sohasida juda noqulay, asosiy vazifalardan biri aniq ilmiy faktlar, masalan, *“kibernetika atamasi nimani anglatadi”* yoki *“ushbu maqolaning muallifi kim”* va axborot tizimining mohiyati shaxslar, faktlar, hujjatlar, nashrlar va boshqalar o'rtasida aloqalarni o'rnatish zarurdir.

Axborotni tizimlashtirishning standart yondashuvi taksonomiyalar yordamida hujjatlarni tasniflashdir. Taksonomiya-bu atamalarni boshqariladigan lug'at (tezaus) sifatida guruhlaydigan va ularni (lug'atlarni) iyerarxik tuzilmalar sifatida tartibga soluvchi predmet tasnifi. Har qanday mavzu sohasini tavsiflash uchun, odatda, ma'lum bir asosiy atamalar to'plami ishlatiladi, ularning har biri ma'lum bir mavzu sohasidagi har qanday tushunchani belgilaydi yoki tavsiflaydi. Tasniflashning asosi tushunchalarni ajratish, ular o'rtasida paradigmatic munosabatlarni o'rnatish (masalan, ota – ona-avlod turi) va tahlil qilingan hujjatni tanlangan tushunchalarga taqqoslashdir.

Ta'lim axborot tizimlarini ta'minlashning eng yoqimsiz tomoni shundaki, so'nggi yuz yil ichida kutubxonalar tomonidan ishlab chiqilgan ma'lumotlarni tasniflash va tizimlashtirish texnologiyalari tasniflangan hujjatlarning tematik yaqinligi tufayli ishlamaydi.

Masalan, UDK va MSC lug'atlari yoki YUNESKOning tezauruslari matematika yoki informatika bo'yicha resurslarni tasniflash uchun eng qulay deb hisoblanadi, qoida tariqasida, ma'lum bir o'quv kursi uchun tanlangan barcha manbalarni bitta toifaga kiritadi[5. 438–475 p]. VINITI tomonidan 1976 yilda nashr etilgan monografiyadan axborot qidiruvining nazariy muammolarini batafsil ko'rib chiqishni o'z ichiga olgan holda, axborot ehtiyojlarining ikki turini – zarur ilmiy ma'lumotlar manbalari to'g'risidagi ma'lumotlarga bo'lgan ehtiyojni va eng zarur ilmiy ma'lumotlarga bo'lgan ehtiyojni ajratib ko'rsatish asosida – birinchisining axborot ehtiyojlarini qondirish uchun hujjatli, ikkinchi turdagi – faktografik deb nomlangan axborot tizimlari mo'ljallanganligi aytiladi. Hozirgi vaqtda har ikkala turdagi axborot tizimlarining imkoniyatlarini birlashtirgan va “hujjat – fakt – tasniflash” sxemasiga muvofiq malakali foydalanuvchining axborot ehtiyojlarini qondirishga imkon beradigan axborot tizimlari ilmiy faoliyatni axborot bilan ta'minlashning eng mashhur vositasiga aylanmoqda. Bunday axborot tizimlarida mantiqiy xulosa chiqarish qoidalarini rasmiylashtiradigan va ma'lum bir hujjatni tavsiflovchi faktlar va tushunchalar bilan ishlaydigan fikrlash tizimi tarkibiy qism sifatida ishlaydi.

Axborot qidirishda tezauruslar.

Xalqaro standartlashtirish tashkiloti (ISO) ta'rifiga ko'ra, tezaurus tushunchalar o'rtasida aniq apriori munosabatlarni o'rnatish uchun rasmiy ravishda tashkil etilgan indeksatsiya tili tomonidan boshqariladigan lug'atdir. Ushbu ta'rif ushbu birliklar orasidagi leksik birliklar va semantik munosabatlarni tezaurusni tashkil etuvchi elementlar sifatida belgilaydi. Tezaurus munosabatlari (tur – tur, qism – butun, kompleks – element, sabab – oqibat) taksonomiya tuzilishiga qo'llaniladi, ya'ni domendagi asosiy taksonomiyalar aniqlanadi.

Tarixiy jihatdan tezauruslar hujjatlarni qo'lda indeksatsiya qilish uchun yaratilgan va ularni yaratishda avtomatik indeksatsiya bilan bog'liq masalalar hisobga olinmagan. Indekslangan ma'lumotlarning barcha tematik xilma-xilligiga mos keladigan tezaurusni qurishdagi qiyinchilik uning zamonaviy axborot tizimlarida mashhur emasligining asosiy sababidir. Ammo, yuqorida aytib o'tganimizdek, ilmiy va o'quv faoliyatini qo'llab-quvvatlash uchun axborot qidirish tizimlarining samaradorligi to'g'ridan-to'g'ri ixtisoslashgan tezauruslardan foydalanishga bog'liq.

Tarixdagi birinchi va bugungi kunda eng mashhur tezauruslardan biri (ideografik lug'atlar) ingliz leksikografi **Piter Mark Rojet** tomonidan tuzilgan va 1852 yilda nashr etilgan [Salton G. 46 p]. Rojet tezaurusining asl nomi (Roget's Thesaurus) Inglizcha so'zlar va iboralarning tezaurusi (Thesaurus of English Words and Phrases). Kompyuterlar bilan bog'liq holda, tezaurus birinchi marta M.Masterman tomonidan 1954 yilda mashinani tarjima qilish sohasida ishlatilgan. Tezauruslardan foydalanib, foydalanuvchi so'rovlari tili va axborot tizimidagi hujjatlar o'rtasida yozishmalar o'rnatildi. Yu.A. Shrader tezaurusni til tomonidan aks ettirilgan bilimlar tizimi sifatida ko'rib chiqishni taklif qildi.

Umumjahon avtomatlashtirilgan tezauruslar orasida RUSSNET rus tiliga o'xshash va yuqorida aytib o'tilgan YUNESKOning ko'p tilli tezaurusiga o'xshash aqlli kompyuter tezaurusi WORDNET (ingliz tilining so'z boyligi) ni ta'kidlash kerak.

Foydalanuvchilar uchun dasturiy ta'minotni qo'llab – quvvatlaydigan mashina vositalarida allaqachon amalga oshirilgan ixtisoslashtirilgan tezauruslar orasida quyidagilarni ta'kidlash mumkin:

- ❖ EuroVOC - Evropa Ittifoqi hujjatlarini indeksatsiya qilish uchun ishlatiladigan katta qidiruv tezaurusi;
- ❖ AQSh Kongressi kutubxonasi tadqiqot xizmatining tezauruslari va boshqariladigan lug'atlari;
- ❖ Amerika ASI indeksatsiya jamiyatining tezauruslari;
- ❖ qishloq va o'rmon xo'jaligi terminologiyasini o'z ichiga olgan taniqli agrovoc tezaurusi;
- ❖ kompyuterlashtirilgan tezaurus tibbiy terminologiya SNOMED va boshqalar.

Ma'lumot tavsifidagi tezauruslar.

Axborot-qidiruv tezaurusi – bu hujjatlar va qidiruv so'rovlarining mazmunini tavsiflash uchun mo'ljallangan, atamalar o'rtasida aniq ko'rsatilgan semantik munosabatlarga ega bo'lgan tabiiy tilda asosiy atamalarining me'yoriy boshqariladigan lug'ati. Tezaurus ma'lum bir domen sohasini tavsiflash uchun mo'ljallangan bo'lib, uning har bir atamasi ma'lum bir domen tushunchasini bildiradi yoki tavsiflaydi.

Tezaurusni tashkil etuvchi atamalar deskriptorlar va askriptorlarga bo'linadi. Tavsiflovchilar domen tushunchalariga aniq mos keladi. Tavsiflovchilar o'rtasidagi munosabatlar odatda ikki turga bo'linadi: **ierarxik** va **assotsiativ**. Ierarxik munosabatlar odatda assimetrik va o'tish davri sifatida qaraladi. Axborot-qidiruv tezauruslarida ishlatiladigan ierarxik munosabatlar alohida turlarga bo'linishi mumkin. Axborot-qidiruv tezauruslarida ishlatiladigan asosiy ierarxik munosabat-bu Rodovid munosabati (ota – ona – avlod, kengroq-torroq, balandroq-pastroq). Agar quyi oqim tavsiflovchisi tushunchasining hajmi yuqori oqim tavsiflovchisi tushunchasining hajmiga kiritilgan bo'lsa, ikkita tavsiflovchi o'rtasida Rodovid aloqasi o'rnatiladi. Shuningdek, ierarxik munosabat sifatida, qism – butun munosabatlar axborot-qidiruv tezauruslarida o'rnatilishi mumkin. Assotsiatsiya munosabati iyerarxik va assotsiativ emas. Axborot qidirish tezaurus

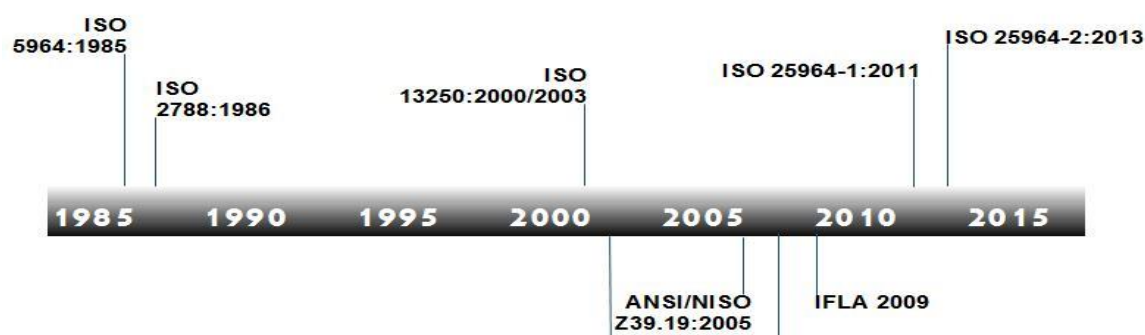
tavsiflovchilari o'rtasida assotsiativ munosabatlarni o'rnatishning asosiy maqsadi indekslash yoki qidirishda foydali bo'lgan qo'shimcha tavsiflovchilarni ko'rsatishdir.

E'tibor bering, milliy va xalqaro standartlarda tavsiflangan axborot-qidiruv tezaurus modeli hujjatlarni qo'lda, ekspert tahlil qilish jarayonida foydalanish uchun mo'ljallangan.

An'anaviy axborot-qidiruv tezauruslarini ishlab chiqishning asosiy maqsadi qo'lda indekslash jarayonida hujjatlarning asosiy mavzularini tavsiflash uchun ularning birliklaridan (tavsiflovchilaridan) foydalanishdir. Shu sababli, axborot qidirish tezaurusining tavsiflovchilari to'plami mavzu sohasidagi hujjatlar mavzusini tavsiflashga imkon berishi muhimdir. Shu bilan birga, bunday tezaurus uchun indeksatsiya jarayonining o'zi lingvistik, grammatik bilimlarga, shuningdek, professional matn indeksatorlari ega bo'lgan fan sohasi haqidagi bilimlarga asoslanadi. Indeksator avval matnni o'qishi, tushunishi va keyin ma'lumot qidirish tezaurusida ko'rsatilgan tavsiflovchilardan foydalangan holda matn tarkibini bayon qilishi kerak. Indeksator matnda ishlatilgan barcha terminologiyani yaxshi tushunishi kerak-matnning asosiy mavzusini tavsiflash uchun unga ancha kam atamalar kerak bo'ladi.

Tezaurusni taqdim etish standartlari.

Tezaurus taqdimotlari uchun bir qator xalqaro (ISO), milliy va korporativ standartlar va tavsiyalar ishlab chiqilgan. ***Quyidagi rasmda ularning evolyutsiyasini aks ettiradi:***



Yangi standart tezauruslar, odatda, ma'lumot topish uchun ishlatiladigan boshqa turdagi lug'atlarga nisbatan qo'llaniladi. U ushbu lug'atlarning elementlari va imkoniyatlarini tavsiflaydi, taqqoslaydi va tahlil qiladi, o'zaro ta'sir o'tkazish zarurati tug'ilganda muhim ahamiyatga ega. Standart bir nechta tezauruslar yoki tezauruslar va boshqa turdagi lug'atlar o'rtasida o'zaro muvofiqlikni o'rnatish va saqlash bo'yicha ko'rsatmalarni o'z ichiga oladi. ISO 25964 standartining ikkala qismi 2005-2007 yillarda nashr etilgan[3. 32 p].

Amaliy tavsiyalar. Lingvistik masalalarni tezkor hal etishda, kompyuter texnologiyalarining ahamiyati beqiyosdir. Ularning milliy tilshunoslikka tatbiq etilishi axborotlarning tezkorlik bilan o'zlashtirilishini, shuningdek bajariladigan amallarning aniqligini ta'minlaydi. Tilshunoslik muammolarining kompyuter yordamida hal qilinishi, ayni paytda, tilning qo'llanish doirasi kengayishi, ichki imkoniyatlarning ortishiga xizmat qiladi.

Xulosa. Hozirgi vaqtda "tezaurus" atamani o'z ichiga oladi va doimiy ravishda yangilanadi. Elektron kutubxonalarda tezaurusdan foydalanish tezaurusni doimiy ravishda modernizatsiya qilish, uning ma'lumotlar bazasiga qo'shilishi va tegishli tematik ixtisoslashuv darajasida eng samarali uslubi hisoblanadi.

Foydalanilgan adabiyotlar:

1. Abduhamidova L. Tilshunoslikning yangi yo‘nalishi: kompyuter lingvistikasi. -Toshkent,- 2015. 75 b.
2. Salton G. Dynamic information and library processing. N. J.: Prentice Hall, 1975. 46 p.
3. ISO 2788:1986. Guidelines for the establishment and development of monolingual thesauri. 2nd ed. Geneva: International Organization for Standardization, 1986. 32 p.
4. Mirahmedova Z. O‘zbek tilining anatomiya terminologiyasi va uni tartibga solish muammolari. – Toshkent: —Fan, 2010 61 b.
5. Masterman M. Semantic message detection for machine translation, using an interlingua // Proc. International Conf. on Machine Translation, 1961. P. 438–475.
6. Лукашевич Н. В. Тезаурусы в задачах информационного поиска. М.: Изд-во МГУ, 2011. 141 с.
7. 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 VI International Conference “Modern Problems of Applied Mathematics and Information Technology - Al-Khorezmiy 2018”, pp. 37–38, Tashkent, Uzbekistan (2018)
8. Abduraxmonova, N. Z. «Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref.» (2018).
9. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
10. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
11. 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.
12. 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;5(03):89-100.
13. 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
14. Abdurakhmonova N. Dependency parsing based on Uzbek Corpus. InProceedings of the International Conference on Language Technologies for All (LT4All) 2019

KOMPYUTER LEKSIKOGRAFIYASI – O‘ZBEK TILSHUNOSLIGINING MUHIM YUTUQLARIDAN BIRI SIFATIDA

Tursunova Sojidxon Axmadjon qizi

O‘zbekiston Milliy universiteti O‘zbek
filologiyasi fakulteti O‘zbek tilshunosligi
kafedrası o‘qituvchisi

Annotatsiya: maqolada lug‘at va uning turlari, an‘anaviy lug‘atlar, elektron lug‘atlar hamda avtomatik lug‘atlar haqida gap boradi. Kitobiy lug‘atlar bilan elektron lug‘atlarning farqlari alohida sanab o‘tiladi. O‘zbek leksikografiyasi va uning tarixi, ilk leksikografiyaning namunasi, Mahmud Koshg‘ariyning “Devoni lug‘otit turk” asari ilk o‘zbek leksikografiyasining namunasi ekanligi etirof etiladi. Korpus leksikografiyasi haqida ham ma‘lumotlar keltiriladi. Brown korpusi, lug‘atlar yaratishda nimalarni hisobga olish kerakligi, L.I.Kolodnyajaya lug‘atshunoslikni qanday turlarga ajratib o‘rganganligi xususida ma‘lumotlar keltirib o‘tiladi.

Kalit so‘zlar: leksikografiya, lug‘at, an‘anaviy lug‘at, elektron lug‘atlar, avtomatik lug‘atlar, kompyuter leksikografiyasi, korpus lingvistikasi va hokazo.

Annotation: the article talks about the dictionary and its types, traditional dictionaries, electronic dictionaries and automatic dictionaries. Differences between book dictionaries and electronic dictionaries are listed separately. Uzbek lexicography and its history, an example of the first lexicography, Mahmud Koshgari's work “Devoni lugotit turk” is recognized as an example of the first Uzbek lexicography. Information about corpus lexicography is also provided. Brown's corpus, what should be taken into account when creating dictionaries, information about how L.I. Kolodnyajaya divided lexicography into types is given.

Keywords: lexicography, dictionary, traditional dictionary, electronic dictionaries, automatic dictionaries, computer lexicography, corpus linguistics, etc.

Аннотация: в статье рассказывается о словаре и его видах, традиционных словарях, электронных словарях и автоматических словарях. Различия между книжными словарями и электронными словарями перечислены отдельно. Узбекская лексикография и ее история, пример первой узбекской лексикографии, примером первой узбекской лексикографии признан труд Махмуда Кошгари «Девони луготит турк». Также представлена информация о корпусной лексикографии. Брауна, что следует учитывать при создании словарей, приведены сведения о том, как Л. И. Колодняжная делила лексикографию на виды.

Ключевые слова: лексикография, словарь, традиционный словарь, электронные словари, автоматические словари, компьютерная лексикография, корпусная лингвистика и др.

Atrofimizni o‘rab turgan butun tabiat doimiy harakatda. Shu bilan bir qatorda ilm-fan sohasida ham doimiy izlanishlar olib borilmoqda. Buning natijasida tilshunoslik sohasida ham yangi yo‘nalishlar paydo bo‘lmoqda, avvaldan mavjud bo‘lganlari esa kun sayin boyib bormoqda. Leksikografiya ham shunday sohalardan biridir. Leksikografiya tilning lug‘at boyligini o‘zida aks ettiradi. O‘zbek leksikografiyasi ham uzoq o‘tmishga ega, uning ilk namunasi sifatida Mahmud

Koshg'ariyning "Devoni lug'otit turk" asari e'tirof etiladi. Asar hijriy 469-yilda yozilgan bo'lib, Abulqosim binni Muhammad Muqtadoga hadya qilingan. Asarda muallif ayni vaqtda iste'molda bo'lgan so'zlarni berganini, boisi bu o'quvchi uchun foydali ekanligini aytadi. Iste'molda bo'lmagan so'zlarni esa kiritmaydi, bu tartib to'g'ri ekanligini ta'kidlaydi.

Leksikografiya tilshunoslik fanining amaliy sohalaridan biri hisoblanadi. U yunoncha so'zdan olingan bo'lib, "lexikos – so'z, lug'at", "grapho - yozaman" degan ma'nolarni anglatadi. Leksikografiya bo'limida lug'atlar, ularni tuzish yo'llari va tamoyillari o'rganiladi.

Avtomatizatsiya ijtimoiy hayotning [barcha jabhalariga](#) kirib kelishi natijasida lug'atlar ham elektron holatda yaratilish imkoniyati yuzaga keldi. Kompyuter yordamida lug'atlar bilan ishlashning optimallasuvi natijasida kompyuter leksikografiyasi yo'nalishi shakllandi.

Kompyuter leksikografiyasi amaliy tilshunoslikning muhim tarkibiy qismi bo'lib, unda lingvistik va dasturiy ta'minot [asosida ishlaydigan](#), kodlash va dekodlash prinsipi asosida yaratilgan kompyuter lug'atlari, ularni tuzish dasturlari, algoritmlari o'rganiladi.

O'tgan davr mobaynida turli xil qomusiy va lingvistik lug'atlar yaratildi hamda takomillashtirib borildi. Texnologiyalar ham kun sayin rivojlandi. Kompyuter texnologiyalari imkoniyatlarining kengayib borishi til bilan bog'liq ishlarning rivojlanishida ham muhim omil bo'lib xizmat qildi.

Leksikografiya sohasiga kompyuter texnologiyalarining tatbiq etilishi lug'at tuzish ishlarini ancha yengillashtiradi. Lug'at tuzish jarayonining naqadar murakkabligini bir tilshunos olim hazilomuz tarzda quyidagicha izohlagan: *"Agar biror kishi og'ir jinoyat sodir qilsa, uni katorgaga surgun qilish shart emas. U lug'at tuzsin, shunda u katorganing barcha azoblarini, [qiyinchiliklarini tortgan bo'ladi](#)".*

Bu borada kompyuter leksikografiyasi (KL) kompyuter lingvistikasining asosiy yo'nalishi sifatida muhim o'ringa ega. Unda kompyuter lug'atlarini yaratish, dasturiy ta'minot uchun lingvistik resurs yaratish vazifalari o'rganiladi. Ayrim manbalarda KL leksikografik manbalarni olish, ular bilan bog'liq lingvistik ma'lumotlarni berilganlar bazasiga yozish, leksik birliklarni kiritish masalalari bilan shug'ullanadi.

L.I.Kolodnyajnyaya lug'atshunoslikni 2 ga ajratadi: an'anaviy hamda raqamli. Tarjima borasida ham ikki yondashuv borligini qayd etiladi: 1) leksikosentrik (kontekstdan kelib chiqib tarjima qilish); 2) slovasentrik (mashina yordamida analiz qilish). [Nilufar Abduraxmonova, 2021: 164].

Kompyuter leksikografiyasi – tilshunoslikning muayyan til leksikonidagi so'zlarning ma'lumotlar bazasi va programmashtirish tillari yordamida ishlash tizimiga ega bo'lgan dasturlar yaratish bilan shug'ullanuvchi sohasi.

Kompyuter va lug'atshunoslik o'rtasidagi o'zaro muvofiqlashuv o'tgan asrning 60-yillarida o'z ifodasini topdi. Bu yillarda Brown korpusi, 1978-yilda avtomatik tekshiruvchi tizim hamda tilni qayta ishlash uchun maxsus kodlangan Zamonaviy ingliz tili Longman lug'ati yaratildi.

Lug'atlarning yaratilishida uch narsa hisobga olinishi kerak:

- lug'atning yaratilishiga ko'ra maqsad va vazifalari;
- lug'atda tavsiflangan matnlar;
- foydalanuvchilarning kategoriyasi. [Nilufar Abduraxmonova, 2021: 164].

Avvalo an'anaviy, elektron va avtomat lug'atlarni bir-biridan farqlash lozim. An'anaviy lug'atlar ensiklopedik va lingvistik turlarga bo'linadi. Elektron lug'atlar an'anaviy lug'atlarga qaraganda foydalanishga oson va qulay. U orqali anchagina vaqtni tejash mumkin. Bu turdagi lug'atlarni bir marotaba yuklab olgandan so'ng, qayta-qayta internetsiz ham foydalanish mumkin.

Shunday bo'lsa-da, avtomatik lug'atlar foydalanuvchiga bir talay imkoniyatlar yaratadi. To'g'ri, ular internet orqali ishlaydi, lekin ularni foydalanuvchilar tomonidan boyitib borish mumkin.

A.N.Baranov tasnifiga ko'ra elektron lug'atlar yaratishning quyidagi bosqichlari bor:

- 1) korpus matnlarini shakllantirish hamda so'zlikni parallel tarzda yaratish;
- 2) namunalarning korpusini avtomatik shakllantirish;
- 3) maqolalarni yozish;
- 4) maqolalarni ma'lumotlar bazasiga kiritish;
- 5) ma'lumotlar bazasida maqolalarni tekshirib ko'rish;
- 6) ma'lumotlar bazasida matnlarni tekshirib ko'rish;
- 7) lug'atdagi matnlarni natijasini olish hamda maketini shakllantirish;
- 8) lug'atni chop etish. [Nilufar Abduraxmonova, 2021: 169].

Ingliz va rus tilshunosligida kompyuter lug'atlari borasida izlanishlar amalga oshirilgan.

Kitobiy lug'atlar va elektron lug'atlarni farqi quyidagilarda ko'rinadi:

1. Kitobiy lug'atlar kartotekalar asosida tuziladi. Elektron lug'atlar korpus (elektron kartotekalar) asosida yaratiladi.
2. Kitobiy lug'atlar qo'lyozma holatda, kompyuter lug'atlari elektron variantda tuziladi.
3. Kitobiy lug'atlar qo'lyozma holida tahrir qilinadi; elektron lug'atlar nisbatan tayyor lingvistik baza asosida yaratiladi.
4. Kitobiy lug'atlar bosma holida nashr qilinadi; elektron lug'atlarda lingvistik ta'minot dasturiy ta'minot bilan uyg'unlashtiriladi.
5. Kitobiy lug'atlar tuzilishi sahifalar ketma-ketligiga tayanadigan chiziqlik tamoyiliga bo'ysunadi. Elektron lug'atlar strukturasi gipertekst texnologiyasiga asoslangan bo'ladi, bu esa foydalanuvchiga lug'at maqolalarining ixtiyoriy qismiga tezkor murojaat qilish imkonini beradi.

Elektron lug'atlarda har bir so'zga mutanaosib keluvchi kod ishlab chiqiladi va qo'llanadi, kodni qayta ishlash jarayonida zaruriy bo'lgan ma'lumotlar, tarjimalar, sinonim, antonym va sharhlarga ega bo'lish mumkin. [Zulxumor Xolmanova, 2020: 76].

Kompyuter leksikografiyasi bugungi kunda turli qo'shimcha imkoniyatlar asosida rivojlanib bormoqda. Xususan, dastlab kompyuter lug'atlari bir tilli, ikki tilli bo'lgan, hozirda ko'p tilli elektron lug'atlar vujudga kelmoqda. Qolaversa, keying vaqtlarda ma'lum bir sohalarga ixtisoslashgan elektron lug'atlar ham yaratilmoqda. Ya'ni, dastlab, kompyuter lug'atlari faqat umumiy leksikon bilangina cheklangan edi, hozirda fanga oid bo'lgan, qurilish, aviatsiya, harbiy, diniy, yuridik sohalarga oid so'zlar bazasi ham e'tiborga olinmoqda. Hatto so'zlarning matnda qo'llanish imkoniyatlari, birikma holdagi ko'rinishlari ham nazarda tutilgan holda lug'atlar yaratilmoqda. Bu ko'rsatkichlar kompyuter leksikografiyasi sohasining taraqqiy etayotganidan darak beradi.

Kompyuter leksikografiyasini elektron matnlar korpusi yoki parallel matnlar korpuslarisiz tasavvur qilish mumkin emas. Matnlar korpusi ("corpus" lotincha "tana" degan ma'noni anglatadi) – bu elektron holda saqlanadigan ma'lum til birliklari bo'lib, ular tilshunoslar uchun turli xil muammolarni hal etish uchun tatbiq etishda va turli yo'nalishdagi tadqiqotlar uchun zaruriyatga qarab turli shakllarda tuziladi. Bular fonema, grafema, morfemalardan tortib undan kattaroq birliklar – leksema, gap va matnlardan (badiiy yoki ilmiy asar, gazeta va jurnal matnlari) tashkil topishi mumkin. Ularning qay tarzda saqlanishiga qarab maxsus dasturlar yordamida har bir kerakli so'z yoki so'z birikmasi uchun darhol uning qo'llanishi bo'yicha misollar topilishi, imlo

bo'yicha variantlari, sinonimik qatorlari topilishi mumkin. Matnlar korpusiga oid ilmiy tadqiqotlar salmog'ining ko'payishi natijasida tilshunoslikda korpus lingvistikasi yo'nalishi shakllandi. [Azamatjon Rahimov, 2011: 85].

Xulosa qilib aytganda, bugungi kunda fan-texnika sohasi jadal rivojlanib bormoqda. Yoshlarning chet tiliga bo'lgan qiziqishlari kun sayin ortmoqda, natijada tilimiz yangi so'zlar va terminlar bilan boyib bormoqda. Bularning barchasi kompyuter leksikografiyasining oldiga yangi-yangi vazifalarni qo'yimoqda.

Foydalanilgan adabiyotlar:

1. Abduraxmonova N. Kompyuter lingvistikasi (darslik), Toshkent: Nodirabegim, 2021. – B. 164
2. Xolmanova Z. Kompyuter lingvistikasi (o'quv qo'llanma), Toshkent: 2020. – B. 64
3. Rahimov A. Kompyuter lingvistikasi asoslari (darslik), 2011. – B. 85
4. www.library.uz
5. www.arxiv.uz

RUS TILSHUNOSLIGIDA ALLUZIYA: O'RGANILISHI VA TURLARI

Karimova Mastura

Mirzo Ulug'bek nomidagi

O'zbekiston Milliy universiteti

O'zbek tilshunosligi kafedrası

E-mail: karimovam94@gmail.com

Annotatsiya: Maqolada alluziya tushunchasining rus tilshunosligida vujudga kelishi, unga berilgan ta'riflar, yondashuvlar haqida fikrlar umimlashtirilgan. Atamaning turli tomondan yoritilishi bilan birga uning turlarga tasniflanishi bilan bog'liq turlicha qarashlar ham keltirilgan.

Kalit so'zlar: intertekstuallik, alluziya, shaxsiy alluziya, metaforik alluziya, yashirin alluziya, strukturaviy alluziya

Аннотация: В статье обобщены размышления о появлении понятия аллюзии в русском языкознании, его определениях и подходах. Наряду с трактовкой термина с разных сторон представлены разные мнения относительно его классификации на виды.

Ключевые слова: интертекстуальность, аллюзия, личностная аллюзия, метафорическая аллюзия, скрытая аллюзия, структурная аллюзия

Annotation: The article summarizes reflections on the emergence of the concept of allusion in Russian linguistics, its definitions and approaches. Along with the interpretation of the term from different sides, different opinions are presented regarding its classification into species.

Keywords: intertextuality, allusion, personal allusion, metaphorical allusion, hidden allusion, structural allusion

Intertekstuallikning bir ko'rinishi sifatida tahlil qilinuvchi alluziv birliklar bugungi kunga kelib badiiy mant tahlilida asosiy ahamiyat kasb etadi. "Alluziya (lot. *Allucio*— *ishora, hazil*) barchaga tanish deb hisoblangan real siyosiy, maishiy, tarixiy yoki adabiy faktga ishora qilishga asoslangan stilistik usul [Yo'ldoshev, 2008]. Ushbu atamaning kelib chiqishi XVI asr o'rtalariga to'g'ri keladi. G. Blum "...biror narsani aniq eslatmasdan, esga solish uchun mo'ljallangan ibora" deya ta'kidlab, alluziyani "*ishora*" so'zi bilan sinonim sifatida qo'llagan [Brower, 1986]. XX asrga kelib alluziya termini bilan bog'liq turli qarashlar paydo bo'lib, qizg'in ilmiy tortishuvlarga sabab bo'ldi. Tilshunos olimlaridan E.Y. Abramova, O.S. Axmanova, I.V. Gubbenet, E.M. Dronova, A.S. Evseeva, M.I. Kiose, A.G. Mamaeva, L.A. Mashkova, I.G. Potylitsyna, E.V. Rosen, I.N. Sofronova, Z. Porat va boshqa tilshunos olimlar bu borada o'z fikrlarini bildirganlar [Никашина, 2016]. Masalan, O.E. Naumova, birinchi navbatda, presedent matnlarning kognitiv, pragmatik, lingvistik va madaniy jihatlariga e'tibor qaratadi. M. Dronovanning ta'kidlashicha, alluziv birliklarni o'rganish madaniyatlararo munosabatlarni, intertekstuallik muammolarini o'rganish doirasida olib borilishi, stilistika, frazeologiya, tarjimashunoslik sohalariga taalluqli bo'lishi mumkin [Dronova, 2004]. A.S. Evseev bu masalani poetika, ritorika, tilshunoslik va madaniyatshunoslik, adabiyotshunoslik kabi fanlarning rivojlanishi bilan bog'liq muammolarga markaziy o'rinlardan biri sifatida baholaydi [Evseev, 1990]. M.I. Kiose alluziyaga ham stilistik vosita, ham uning natijasi sifatida qaraydi: "...Alluziya (ishora) usuli ancha keng tarqalgan bo'lib,

turli janrdagi obrazga boy matnlarni yaratishda, ularning emotsional-bahoviy mazmunini oshirishda katta ahamiyatga ega”[Kiose, 2002]. Bunda alluziyaning asosiy vazifalari haqida keltiriladi: asar yaratilgan vaqni o‘rganish, qahramonlar xarakterini tavsiflash, matnga milliy “rang” berish, muallif fikrini retsipientga yetkazish. L.A. Mashkova “...alluziya – ikkita matnni birlashtiradi, bunda matn retsipientga aniq tasavvur uchun ko‘maklashadi va o‘zidan oldingi yaratilgan asarlar haqida ham ma‘lumot beradi” degan fikrlarni bildiradi[Mashkova, 1989]. Yana bir tilshunos V.E. Xalizev alluziyaning tavsiflashda intertekstlar orasidagi munosabatga e‘tibor qaratadi: “Alluziya tarixiy o‘tmish haqidagi asarlarda zamonaviy ijtimoiy hayot voqeligiga ishoradir” [Никашина, 2016]. Tadqiqotchi L. I. Lebedeva ushbu atamaga batafsilroq to‘xtalgan : “ ... alluziya (lot. hazil) – uslubiy vosita sifatida taniqli narsalarga ishora. Nutqda, adabiy asarda, ilmiy ishda va boshqalarda umumlashgan subtekstni yaratadigan tarixiy, afsonaviy, kundalik fakt” [Никашина, 2016]. Alluziya tilshunoslik va adabiyotshunoslik kesishmasida paydo bo‘lgan tushuncha. Bu borada adabiyotshunoslar fikri ham deyarli bir xil.

Alluziya tushunchasi haqida bir to‘xtamga kelingandan so‘ng, uni sinflarga, turlarga tasniflash ham keyingi masala hisoblandi. Ingliz tilidagi adabiy atama lug‘atlariga ko‘ra, alluziya turlarining eng to‘liq tasnifi quyidagilardan iborat:

- 1) **alluziya** – yaqinda sodir bo‘lgan voqealarga ishora (topik ishora);
- 2) **shaxsiy alluziya** - yozuvchining tarjimai holidagi faktlarga havolalar (shaxsiy ishora);
- 3) **metaforik alluziya** - uning maqsadi tegishli ma‘lumotlarni yetkazishdir;
- 4) boshqa yozuvchilar uslubiga taqlid qiluvchi **yashirin alluziya** (taqlidli ishora);
- 5) boshqa asarning tuzilishini aks ettiruvchi **strukturaviy alluziya** (strukturali ishora) [Ben-Porat,1976].

Rus tilshunoslaridan birinchilardan bo‘lib Olima Novoxacheva 2 xil alluziyaning ajratadi: intertekstga o‘tish jarayonida o‘zgarishga uchraganlar va o‘zgarishsiz berilganlar. Keyinchalik alluziyaning bir qator turlari tasniflana boshlandi.

1. Leksik- semantik alluziya
2. Stilistik alluziya
3. Morfologik alluziya

Yuqoridagi 3 turga ajratish alluziv birliklarning barchasini qamrab olmaganligi sababli yana bir tilshunos olim M.D. Tuxareli alluziyaning 3 turini ajratib ko‘rsatadi:

- 1) antroponimlar – o‘ziga xos ismlar, kosmonimlar – sayyoralar, yulduzlar nomlari, toponimlar – geografik nomlar, zoonimlar – hayvonlar, qushlarning nomlari, tematonimlar – tarixiy voqealar, bayramlar va boshqalar nomlari, teonimlar – ma‘budlar, jinlar, mifologik belgilar va boshqalarning nomlari;
- 2) mifologik, bibliyaviy, tarixiy, adabiy va boshqa voqeliklar;
- 3) ishora iqtiboslar, eslatmalar [Ben-Porat,1976].

Alluziya turlari har bir millat, xalq, umuman olganda, yashash tarzi yagona bo‘lgan insonlar uchun turlicha talqin qilinadi. Keltirilgan fikrlar yana vaqt o‘tishi bilan yangicha tus olishi hech ajablanarli hodisa emas. Chunki adabiyot yangilanar ekan, uning tahlili ham yangilanib boraveradi. Rus tilshunoslarining alluziya borasidagi fikr va mulohazalari tahlili shuni ko‘rsatadiki, o‘rganilayotgan hodisa ko‘p qirrali va yangi jihatlarini ilmiy izlanishlarda ochib berilmoqda.

Foydalanilgan adabiyotlar:

1. Yoʻldoshev M. Badiiy matnning lisoniy tahlili. Oʻquv qoʻllanma. – T., 2008. – B. 120.
2. Dronova E.M. (2004). Problems of translation of the stylistic device of allusion in Anglo-Irish literature of the first half of the twentieth century // Bulletin of VSU. No. 1. pp. 83-86.
3. Evseev A.S. (1990). Fundamentals of the theory of allusion (based on the material of the Russian language) spec.
4. Kiose M.I. (2002). Linguistic-cognitive aspects of allusions: Based on the headings of English and Russian journal articles: Candidate of Philology: 02/10/20. Moscow.
5. Mashkova L.A. (1989). Allusivity as a category of vertical context [Allusivity as a category of vertical context] // Bulletin of the Moscow University. Series 9. Philology. No. 2. pp. 25-33.
6. Ben-Porat Z. (1976). The Poetics of Literary Allusion. PTL: A Journal for Descriptive Poetics and Theory of Literature 1. P. 105—128.
7. Brower R.A. (1986). Alexander Pope: The Poetry of Allusion. Oxford: Oxford at the Clarendon Press.
8. Galperin I.R. (1977). Stylistics. M.: Higher School Publishing House.
9. Н.В. Никашина, Н.Д. Супрун. Аллюзия как стилистический прием в англоязычной литературе . Вестник РУДН, серия Теория языка. Семиотика. Семантика, 2016, № 1
10. Ben-Porat Z. (1976). The Poetics of Literary Allusion. PTL: A Journal for Descriptive Poetics and Theory of Literature 1. P. 105—128

BILINGVAL TEZAVURUS YARATISHDA SOHAVIY TERMINOLOGIYA MASALALARI

Oybekova Sarvinoz

O'zbekiston Milliy universiteti magistranti

Annotatsiya: Mazkur maqolada terminologiya sohasi va uning zamonaviy tilshunoslik nuqtayi nazaridan tahlil etish haqida bahs yuritiladi. Xususan, terminologik lug'atlarni yaratishda ularni sohalarga ko'ra ajratish tamoyillari hamda ularni bir tilli, ko'p tilli va ikki tilli shaklda ishlab chiqish haqida so'z yuritiladi. Maqola obyekti esa bilingval tezavrus-lug'atlardir. Bunda terminologiyaga oid ikki tilli lug'at-tezavruslar yaratishning nazariy masalalari tahlilga tortiladi.

Kalit so'zlar: lug'at, tezavrus, bir tilli lug'at-tezavrus, ko'p tilli lug'at-tezavrus, ikki tilli (bilingval) lug'at-tezavrus, atama, terminologiya.

Abstract: This article discusses the field of terminology and its analysis from the point of view of modern linguistics. In particular, when creating terminological dictionaries, the principles of dividing them according to fields and their development in monolingual, multilingual and bilingual form are discussed. The object of the article is a linguistic thesaurus-dictionaries. In this, the theoretical issues of creating bilingual dictionaries-thesauruses related to terminology are analyzed.

Keywords: dictionary, thesaurus, monolingual dictionary-thesaurus, multilingual dictionary-thesaurus, bilingual (bilingual) dictionary-thesaurus, terms, terminology.

Аннотация: В данной статье рассматривается область терминологии и ее анализ с точки зрения современной лингвистики. В частности, при создании терминологических словарей обсуждаются принципы их деления по отраслям и их развитие в одноязычной, многоязычной и двуязычной форме. Объектом статьи являются лингвистические тезаурусы-словари. При этом анализируются теоретические вопросы создания двуязычных словарей-тезаурусов, связанных с терминологией.

Ключевые слова: словарь, тезаурус, одноязычный словарь-тезаурус, многоязычный словарь-тезаурус, двуязычный (двуязычный) словарь-тезаурус, термины, терминология.

Til nutq birliklari, xususan, so'zlar yordamida insonlar bilan muloqot o'rnatish vositasidir. So'zlar ana shunday aloqa quroli vazifasini bajarish jarayonida ma'lum bir tasniflanishlarga uchraydi. Masalan, ulardan biri biror soha yoki fanga oid atamalar tizimi – terminologiyadir. Terminologiya biror bir sohani tushunish va unga oid mavzular haqida muloqot qilish imkoniyatini yaratadi. Masalan, iqtisod haqida gapirish, suhbat jarayonida maxsus terminlarni qo'llash uchun kontekstning ma'lum talablariga bo'ysunishga majbur bo'lamiz.

Terminologiya resurslarini nutq jarayonida qo'llash kundalik hayotimizdan boshlanib, yuqori ilmiy yoki kasbiy saviyagacha ko'tarilishi mumkin. Biroq soha mutaxassisi bo'lmaganlar uchun bunday atamalardan foydalanish variant bo'lsa, mutaxassislar uchun ilmiy yoki kasbiy ehtiyoj hisoblanadi. Terminologiya ma'lum bir yo'nalishga xos bo'lgan bilimlarni amaliyotga qo'llash, muloqot qilish va o'rgatish bilan shug'ullanadiganlar uchun muhim zaruratdir. Qaysi sohada bo'lmasin mutaxassislar o'z fikrlarini ifodalash uchun terminologiya imkoniyatlaridan foydalanadilar. Terminologiya haqida yetarli tushunchaga ega bo'lish barcha sohalarga kirishning

ilk bosqichi sanaladi. Terminologiyani bilmasdan faoliyat boshlansa, mutaxassis hech qachon u sohaning ichiga kirolmaydi, faqat yuzaki xulosalar bilan cheklanishga majbur bo'ladi.

“Terminologiya bilan nafaqat ilmiy-texnik mutaxassislar, balki ilm-fanni ommalashtirish va uning targ'ibotiga chog'langan barcha kommunikativ vositachilar, ixtisoslashgan tarjimonlar hamda rahbarlar, texnik yozuvchilar, til o'qituvchilari muqarrar ravishda terminologiyaga muhtojlik sezadilar¹⁸”.

Kunimizda barcha hujjatlar terminologik tilda yozilgan bo'lib, ularning katta qismini noprofessionallar tushunmaydi. Terminologiya aniq sohaga oid masalalarni batafsil tushunishga, turli soha vakillariga yanada samaraliroq ma'lumot almashinishga yordam beradi.

Bir tili, ikki tili yoki ko'p tili terminologik ma'lumotlar bazasini ishlab chiqish bevosita terminologiyani boshqarish bilan bog'liq. Terminologiyani boshqarish deganda tarjima jarayonining muhim bir bosqichini tushunamiz. Masalan, tarjimonlar ilmiy yoki sohaviy materiallarni tarjima qilish jarayonida yanada aniqlik va izchillikka erishish uchun terminologiya yutuqlaridan foydalanadilar. Tarjima ma'lum matnni bir tildan boshqa tilga ko'chirib qo'yish emas, turli madaniyatlarga taalluqli kishilar ma'lum tushunchani qanday ifodalashini ko'rsatib berishni o'z ichiga oladi¹⁹.

Tezaurus tildagi yoki ma'lum sohaga oid barcha so'zlarni qamrab oladigan, ularning kontektda qo'llanilish holatlarini to'liq o'zida aks ettiradigan lug'atdir. Terminologiya oid lug'at-tezaurus lug'atlarning boshqa turlaridan bir sohaning “hajmli” kognitiv modelini o'zida jamlashi bilan ajralib turadi. Tezaurusning strukturaviy parametrlarini terminlarni tizimlashtirish usuli sifatida kontseptual tasniflash tashkil etadi²⁰. Yaxshi shakllangan tezaurus tarkibida bir-biriga o'zaro bog'liq bo'lgan bir necha qismlar mavjud bo'ladi. Tasniflash sxemasi ma'lum sohaga oid terminlarning barchasini katta kontseptual zonalarga bo'ladi va ularni o'ziga xos tizimlar asosida boshqaradi.

Tezauruslarda eng ko'p qo'llaniladigan atamadan keyin alifbo tartibida, atama elementi sifatida ushbu terminni o'ziga qamrab olgan barcha tezaurus atamalar keltiriladi va har bir so'z o'rin almashish indeksida nechta keng qo'llaniladigan atamani o'ziga jamlagan bo'lsa, shu miqdorda takrorlanadi.

Ikki tili lug'at yoki tarjima lug'ati so'z yoki iboralarni bir tildan boshqa tilga tarjima qilish uchun ishlatiladigan maxsus lug'atdir. Ikki tili lug'atlar bir yo'nalishli bo'lishi mumkin, ya'ni ular bir tildagi so'zlarning ma'nolarini boshqa tilda sanab o'tadi yoki ikki tomonlama bo'lishi mumkin, bu ikkala tilga va ikkala tildan tarjima qilish imkonini beradi. Ikki tomonlama ikki tili lug'atlar odatda ikkita bo'limdan iborat bo'lib, ularning har birida bir tildagi so'z va iboralar alifbo tartibida hamda tarjimasi bilan birga ko'rsatiladi. Tarjimaga qo'shimcha ravishda, ikki tili lug'at odatda nutq qismini, jinsini, fe'l turini, declension modelini va boshqa grammatik maslahatlarni ona tili bo'lmagan kishiga so'zdan foydalanishga yordam beradi. Ba'zan ikki tili lug'atlarda mavjud bo'lgan boshqa xususiyatlar - iboralar ro'yxati, foydalanish va uslublar bo'yicha qo'llanmalar, fe'l jadvallari, xaritalar va grammatik havolalar. Ikki tili lug'atdan farqli o'laroq, bir tili lug'at so'z va iboralarni tarjima qilish o'rniga ularni belgilaydi.

¹⁸ Cabré, M. Teresa (2002). “Terminología y normalización lingüística”. Jornadas (EHU: LEIOA) Terminología y lenguajes de especialidad. Euskara Institutua EHU-LEIOAKO CAMPUSA País Basc.

¹⁹ Mayra Leon. [<https://leontechtrans.com/importance-of-terminology-and-terminology-management>]

²⁰ Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. М.: Наука, 1981.

Ikki tilli terminologik lugʻat-tezauruslar platformasini yaratishni oʻrganar ekanmiz, hozirda mavjud boʻlgan elektron lugʻatlar va platformalar haqida ham maʼlumotga ega boʻlishimiz kerak. Bugungi kunda kasbiy faoliyat va taʼlim sohasida axborot texnologiyalarining keng qoʻllanilishi bilan elektron resurs, jumladan, elektron lugʻatlardan keng foydalanilmoqda. Elektron lugʻatlar yaratish tarixi oʻtgan asr oʻrtalariga borib taqaladi. Kompyuter yordamida yaratilgan birinchi leksikografik manba Georgetown University Machine Translation Research and Language Project jarayonida 1956-yilda jamlangan Jorjtaun universiteti lugʻat bazasi edi. Soʻngra 1963-yilda Vebsterning 7-kompyuter nashri, 1966-yilda “Random House” va 1969-yilda “American Heritage” nashriyotlari izohli lugʻatlarining elektron nashrlari paydo boʻldi. Shu oʻrinda 1978-yilda Longman, 1979-yilda Harper Collins kabi ingliz tilidagi izohli lugʻatlarning elektron versiyalarini ham keltirib oʻtishimiz kerak²¹.

Bugungi kunda keng qoʻllaniluvchi ikki tilli lugʻatlar sifatida quyidagilarni koʻrsatishimiz mumkin:

1. Lingvo (X3, X5 versiyalari).
2. Multitran.
3. Polyglossum (3.52, 3.72 versiyalari).
4. MultiLex 6 (“Compact” va “Deluxe” versiyalari).

Hozirda sohaviy bir tilli elektron lugʻatlar ham faol qoʻllanilmoqda. Masalan, “Fan va texnologiyalarning akademik matbuot lugʻati”, “Onlayn buxgalteriya lugʻati”, “Sanʼat lugʻati”, “Bank va moliya lugʻati”, “Qonun lugʻati”, “Onlayn tibbiy lugʻat” kabi elektron lugʻatlar shular sirasindandir. D.I.Mendeleyevning onlayn jadvali, kimyoviy elementlarning talaffuzi va nomlarini 7 tilda koʻrsatuvchi “WebElements” lugʻatini koʻp tilli tarjima elektron sohaviy lugʻat sifatida ayrica ahamiyat kasb etadi.

Terminologik lugʻatlarda keng koʻlamli predmet sohalari terminlarini izohlash muhimdir. Bunda axborot-qidiruv tezauruslarini yaratish tajribasiga suyanish kerak. Tajribadan koʻringani kabi u quyidagi asosiy bosqichlardan iboratdir:

- axborot-qidiruv konteksti eʼtiborga olinadi;
- resurs birliklari atamalar bildirgan tushuncha zahirida yaratiladi;
- katta miqdordagi koʻp soʻzli (tarkibli) ifodalar tahlil qilinadi, bunda koʻp soʻzli birliklarni kiritish (kiritmaslik) tamoyillari biriktiriladi;
- birliklar tushunchalari orasidagi munosabatning kam sonli toʻplami vujudga keltiriladi.

N. V. Lukashevichning fikriga koʻra, matnlarni avtomatik qayta ishlashda lingvistik bazadan foydalanish haqida soʻz yuritilar ekan, bunda WordNet turiga oid leksik resurslarni yaratish metodologiyasidan foydalanish maqsadga muvofiq. Bunda quyidagi qoidalarga amal qilamiz:

– maʼlum tushunchaga ega boʻlgan birliklar aynan mavjud boʻlgan til birliklari maʼnolari asosida yaratiladi;

– leksik-terminologik tushunchalar tizimining koʻp pogʻonali iyerarxik tuzilishi ishlab chiqiladi;

- koʻp maʼnoli soʻz va ifodalarning maʼnolarini tavsiflash tamoyillari oʻrnatiladi²².

²¹ Карпов Ю.М., Лейчик В.М. Применение методов кластер-анализа для решения терминологических задач // Проблемы информационных систем. – М., 1986 - №91. – С. 69-87.

²² Abjalova M. Oʻzbek tili ontologiyasi: yaratish texnologiyasi va konsepsiyasi. Monografiya. – Toshkent: Nodirabegim, 2022. – 83 b.

Terminologiya har bir soha uchun asosiy qurilish blokidir. Tarjima va mahalliyashtirish jarayonida hal qiluvchi rol o'ynaydi. Terminologiya hujjatlaringizning aniqroq va izchil mahalliyashtirilgan versiyasini yaratishga yordam beradi. Terminologiya umumxalq tilining leksikasi negizida vujudga keladi, boyiydi. O'z navbatida u, umumxalq tilining boyishi uchun xizmat qiladi. Terminologik leksika umumxalq tili leksikasining ajralmas qismi sifatida u bilan birga rivojlanadi, jamiyat va til taraqqiyotidagi barcha jarayonlarni boshdan kechiradi. Ma'lum bir soha yoki fan yuqori darajada taraqqiy etgandagina o'sha tilda maxsus terminologiya vujudga keladi.

Foydalanilgan adabiyotlar ro'yxati:

1. Cabré, M. Teresa (2002). "Terminología y normalización lingüística". Jornadas (EHU: LEIOA) Terminología y lenguajes de especialidad. Euskara Institutua EHU-LEIOAKO CAMPUSA País Basc.
2. Mayra Leon. [<https://leontechtrans.com/importance-of-terminology-andterminology-management>]
3. Дубичинский В.В. Теоретическая и практическая лексикография. Вена Харьков, 1998.
4. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. М.: Наука, 1981.
5. Кобрин Р.Ю., Антонова М.В. Терминологические системы и их когнитивные модели // Очерки научно-технической лексикографии. СПб.: Изд-во С.-Петерб. ун-та, 2002. С. 47 – 66.
6. Карпов Ю.М., Лейчик В.М. Применение методов кластер-анализа для решения терминологических задач // Проблемы информационных систем. – М., 1986 - N91. – С. 69-87.

RUSCHA-O'ZBEKCHA MASHINA TARJIMASIDA MATN TAHLILI MASALALARI

Murodova Dildora

O'zbekiston Milliy universiteti 3-kurs talabasi

E-mail: murodovadildora@gmail.com

Annotatsiya: Ushbu maqolada korpus texnologiyasi yordamida ruscha-o'zbekcha mashina tarjimasidagi matn tahlili masalalarining o'rganilishi va olimlarning mavzuga oid asarlari to'g'risidagi fikrlari yoritiladi hamda ularga munosabat bildiriladi. Shuningdek, rus tilidan o'zbek tiliga tarjima qilishdagi o'zgarishlar qiyoslab beriladi.

Kalit so'zlar: Mashina tarjimasi, o'zbek tili, rus tili, kompyuter lingvistikasi, parallel korpus matnlari, mashina tarjimasida morfologik tahlil.

Annotation: This article examines text analysis issues in Russian-Uzbek machine translation using corpus technology and the opinions of scientists about the works on the subject are highlighted and they are treated. Correspondingly, the changes in translation from Russian to Uzbek are compared.

Keywords: machine translation, Uzbek language, Russian language, computer linguistics, parallel corpus texts, morphological analysis in machine translation.

Аннотация: В данной статье с помощью корпусной технологии исследуются проблемы анализа текста в русско-узбекском машинном переводе и высказываются мнения ученых о работах по теме и их реакции. Также сравниваются изменения в переводе с русского на узбекский.

Ключевые слова: машинный перевод, узбекский язык, русский язык, компьютерная лингвистика параллельные корпусные тексты, морфологический анализ в машинном переводе

Bugungi kunda ijtimoiy-siyosiy hayotimizda tilning ahamiyati barcha sohalarda o'z aksini topmoqda. XIX asrning oxiri XX asrning boshlarida kompyuter texnologiyalarning rivojlanishi va internet tarmog'ining o'sishi natijasida tilshunoslikda kompyuter lingvistikasi va korpus tilshunosligi sohalarining rivojlanishi uchun keng imkoniyatlar yaratildi". XVIII asrda faylasuflar G.V.Leybnis va Deskard tomonidan jumla va so'zlarning o'zaro bog'liqligini kodlash masalasi ilgari surilgan. Mashinaning tarjima qila olish imkoniyati haqidagi g'oyani CH.Bebbidj (1791-1871) 1836-1848- yillarda o'sha davrdan 100-yil keyin paydo bo'lgan elektron raqamli mashinalarning mexanik prototipi- raqamli analitik mashinalar bo'yicha olib borgan loyihasida qayd etilgan. 1933-yilda Rossiyada lingvistik arifmometr muallifi P.P.Smironov Troyanskiy tomonidan esperanto tili asosida tillar o'rtasidagi grammatik boshqaruv usuli bilan taqsimlash metodi va bilingv lug'atni o'z ichiga oluvchi tizim yaratildi". Olima N.Abdurahmonovanning mashina tarjimasining shakllanishi uning taraqqiyot bosqichlari, dastlabki mashina tarjimasining paydo bo'lish evolyutsiyasi haqidagi ilmiy izlanishlari diqqatga sazovor [Abdurahmonova N.: 2012].

"Biz statistik mashina tarjimasi sifatini yaxshilash uchun yuqori darajada assimetrik morfologik tuzilmalarga ega bo'lgan ikki til o'rtasidagi morfologik va sintaktik simmetriyani

keltirib chiqaradigan yangi morfologik tahlil usulini ko'rishimiz mumkin. Bunda morfologik jihatdan boy tildagi so'z prefiks(lar) o'zak qo'shimcha(lar)i qatoriga va parallel korpusning nutq belgilarining bir qismiga ega hisoblanadi. Arab (teg) va ingliz (teg) o'rtasidagi eng muhim bo'lgan nutq qismi yozishmalarni aniqlaydigan tegdan tegga tarjima qilish ehtimolini olamiz". Bunda arab tilidan ingliz tiliga tarjima qilishda morfologik tahlil vazifasidagi morfemalarning ma'noli o'zaklarga birlashishi yoki o'chirilishini ifodalaydi [Young-Suk Lee. :2004].

"Matnni kompyuter yordamida tarjima qilishda tahlil masalalari muhim sanaladi. Morfologik tahlilda so'z turkumlari va ularning grammatik kategoriyalarining modellarini ingliz va o'zbek tillarida fe'l va ot so'z turkumi misolida tahlil qilinadi. Kompyuter har bir so'zni morfologik darajada tahlil qila olishi undagi ma'lumotlar bazasining to'liq va to'g'ri ekanligini ko'rsatadi. Sintaktik jihatdan matnni tahlil qilishda N.Chomskiyning formal til grammatikasiga asoslaniladi. Ingliz tilidan o'zbek tiliga tarjima qilishda gaplarni daraxtsimon sintaktik tahlil qilish, ega va kesim guruhidagi tobe munosabatli birikmalarni topishda muhim rol o'ynaydi. Morfologik tahlil so'zlarning so'z turkumlari bazasi, sintaktik tahlil so'z birikmasi lug'ati va gap qoliplari bazasiga murojaat qiladi"[Abdurakhmonova and Teshaboyev.: 2019].

Rus tili milliy korpusi 18- asr o'rtalaridan 21- asr boshigacha bo'lgan davriy manbalarni qamrab olgan. Bu korpusda o'tgan va hozirgi davr o'rtasidagi sitsiolingvistik variantlar: adabiy, dialektal, so'zlashuv tiliga xos bo'lgan adabiyotlar. Bundan tashqari, ocherk, esse, jurnalist nashr, ilmiy va ommabop adabiyotlar yozishma va kundalik hujjatlardan iborat.

Korpus tilshunosligi – hisoblash tilshunosligining bir qismi hisoblanadi. Kompyuter texnologiyalardan foydalangan holda lingvistik korpuslarni (matn korpuslarni) qurish va ishlatishning umumiy tamoyillarni ishlab chiqadi. Matnlarning lingvistik ya'ni til korpusi deganda, muayyan til muommolarni hal qilish uchun mo'ljallangan, mashinada o'qiladigan, birlashtirilgan, tuzilgan filologik jihatdan mukammal lingvistik ma'lumotlar to'plami tushuniladi.

Davlat rahbari yoshlarning tashabbuslarini qo'llab-quvvatladi	Глава государства поддержал инициативы молодежи
Prezident Shavkat Mirziyoyev Davlatobod tumanida joylashgan "Yangi O'zbekiston yoshlar ziyo maskani" kutubxonasini borib ko'rdi.	Президент Шавкат Мирзиёев посетил библиотеку «Янги Ўзбекистон ёшлар зиё маскани» в Давлатабадском районе.
Bu majmua 2021-yil foydalanishga topshirilgan edi.	Этот комплекс был введен в эксплуатацию в 2021 году.
Bugungi kunda 70 mingta kitob fondiga ega.	Сегодня его фонд насчитывает 70 тысяч книг.
Kutubxona jahondagi yirik elektron kutubxonalar tarmog'iga ulangan.	Библиотека подключена к крупнейшей в мире сети электронных библиотек.
Bu orqali xorijiy adabiyotlardan ham erkin va bepul foydalanishi mumkin.	Благодаря этому читатели имеют свободный и бесплатный доступ к зарубежной литературе.
Tashrif buyuruvchilar maxsus ID-karta oladi.	Посетителям оформляются специальные ID-карты.
Barcha jarayonlar elektron tarzda bajariladi.	Все процессы осуществляются в электронном виде.
Birinchii qavatda 12 yoshgacha bo'lgan bolalar hamda ko'zi ojiz insonlarga mo'ljallangan alohida o'qish zali tashkil etilgan.	На первом этаже есть отдельный читальный зал для детей до 12 лет и незрячих лиц.

Biz ruscha-o'zbekcha mashina tarjimasidan foydalanishda avvalo 1000 ta fe'lning uchta zamonda tuslab chiqishdan boshladik. Ya'ni har bir zamonda har bir shaxsda tuslanishda qo'shimchalarning o'zgarishini ko'rishimiz mumkin. Namunamizda "поддержаль" fe'li berilgan.

	I	II	III	I	II	III
Hozirgi zamon	Поддерживаю	Поддерживаешь	поддерживать	Поддерживаем	Поддерживаете	поддерживают
O'tgan zamon	поддерживал	поддерживаль	поддеживал	поддерживали	поддерживали	поддерживали
Kelasi zamon	буду поддерживать	будешь поддерживать	будете поддерживать	будем поддерживать	будет поддерживать	Будут поддерживать

Глава государства поддержал инициативы молодежи.

Davlat rahbari yoshlarning tashabbuslarini qo'llab quvvatladi.

Ushbu gap juftliklarida quyidagi leksik birliklarning muqobilligini ko'rish mumkin: *глава-rahbar, государства- davlat, поддержал -qo'llab quvvatladi, инициативы-tashabbuslarni, молодежи- yoshlarning.*

Rus tilidan o'zbek tiliga tarjima qilishda, ikki til o'rtasidagi farqli jihatlarni ko'rishimiz mumkin. Rus tili flektiv tillar oilasiga mansub, o'zbek tili agglyutinativ tillar oilasiga kiradi. Flektiv tillar agglyutinativ tillardan farqli jihati shuki, o'zakka affiks alohida mustaqil qo'shilmaydi, balki o'zakka singib ketadi. Natijada yuqoridagi misolimizda o'zak shakli, tarkibi o'zgargan.. Agglyutinativ tillarda so'zning o'zagini o'zgartirmagan holda qo'shimchalar qo'shish bilan ma'lum so'z va shakl yasaladi. Ya'ni o'zbek tilida asosga qo'shimchalar zanjirsimon ketma ketlikda qo'shib ketadi. Bugungi kunda parallel korpusda matnlarni tarjima qilishdagi farqli tomonlari ustida tadqiqot ishlari amalga oshirilmoqda.

Morfologik tahlil tilga xos qoidalarni qo'llash orqali amalga oshirilishi mumkin. Bular kontekstni ajratib ko'rsatish bilan to'liq miqyosli morfologik tahlilni yoki bunday manbalar mavjud bo'lmaganda, oddiy evristik qoidalarni, masalan, so'zning morfologik qo'shimchasi sifatida so'zning oxirgi bir nechta belgilarni o'z ichiga olishi mumkin.

Foydalanilgan adabiyotlar:

1. Abdurahmonova N. Mashina tarjimasining lingvistik asoslari. - T.: Akademnashr, 2012.
2. Young-Suk Lee IBM T. J. Watson Research Center, Yorktown Heights, NY 10598
Email: ysuklee@us.ibm.com/ Morphological Analysis for Statistical Machine translation, 2004-apps.dtic.mil
3. 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), Tashkent, Uzbekistan, 2021, pp. 1-4, doi: 10.1109/ICISCT52966.2021.9670043.
4. Abduraxmonova N. Z. Kompyuter lingvistikasi (darslik)-Melbrun: Globe edit, 2020
5. N. Abdurakhmonova, I. Alisher and R. Sayfulleyeva, "MorphUz: Morphological Analyzer for the Uzbek Language," 2022 7th International Conference on Computer

Science and Engineering (UBMK), Diyarbakir, Turkey, 2022, pp. 61-66, doi:
10.1109/UBMK55850.2022.9919579.

6. G.Stanovskiy, NA Smith, L.Zettlemoyer-arxiv preprint arxiv:1906.00591,2019-arxiv.org
7. N.Abdurakhmonova- Journal of Social Sciences and...., 2017- journals.researchhub.org
8. <http://ruskorpora.ru>
9. <https://uzbekcorpus.uz/>