

KORPUSGA ASOSLANUVCHI GRAMMATIK VA LEKSIKOGRAFIK TADQIQOTLAR

Rajabova Karomat Chorshanbi qizi

O'zMU, Kompyuter lingvistikasi mutaxassisligi magistranti.

Annotatsiya: Ushbu maqolada korpusga asoslanuvchi grammatik va leksikografik tadqiqotlar, korpus lingvistikasi oldida turgan asosiy vazifalar, lingvistik masalalar o'z aksini topgan. Leksikografiya, ideografik lug'atlarning korpus lingvistikasidagi o'rni va ahamiyati misollar orqali ko'rsatilib, ular haqida mulohazalar bildirilgan.

Kalit so'zlar: korpus lingvistikasi, leksikografiya, ideografik lug'at, ideografik teglash, tezaurus

Annotation: Grammatical and lexicographic studies based on the corpus, the main tasks and linguistic issues facing the corpus linguistics are reflected in this article.

Keywords: corpus linguistics, lexicography, ideographic dictionary, ideographic tagging, thesaurus

Аннотация: В статье отражены грамматические и лексикографические исследования на основе корпуса, основные задачи и лингвистические проблемы, стоящие перед корпусной лингвистикой.

Ключевые слова: корпусная лингвистика, лексикография, идеографический словарь, идеографическая разметка, тезаурус

Bugungi kunda til korpuslarida turli xil tadqiqotlar olib borilmoqda. Ular orasidan leksikografik tadqiqotlarni alohida ko'rsatish o'rinlidir. Buyuk lug'atshunoslardan biri V.I.Dal shunday yozadi: "Алфавитный способ расположения слов крайне туп и сух. Самые близкие и сродные речения разносятся далеко врозь и томятся тут и там в одиночестве; всякая живая связь речи разорвана и утрачена, читать такой словарь нет сил, на десятом слове ум притупеет и голова вскружится, потому что ум наш требует во всем какой-нибудь разумной связи, постепенности и последовательности" [С.Баранов.1995]. Ya'ni, V.I.Dalning fikricha, so'zning lug'atda joylashish tamoyili lug'atning o'quvchi uchun foydalanish qulayligi bilan belgilanadi.

Til ongimizda sistema sifatida yashashini bilamiz. N.V. Krushevskiy so'zlarni lug'atda joylashtirish masalasiga to'xtalar ekan: "Hech qachon tilning asosiy xarakterini unutmaslik lozim. So'z-narsaning belgisi. So'z, narsa haqidagi tasavvur assotsiatsiya haqidagi qonun asosida juftlik hosil qilgan. Shunday ekan, so'zlar bizning ongimizga ham o'zi anglatadigan jismlar guruhi sifatida kirishi kerak" [Крушевский Н. 1998: 296], – degan edi. Leksikani ideografik nuqtayi nazardan o'rganishning asosiy maqsadi so'zning turli ma'nolarini bus-butun, yaxlitligicha o'rganish, ularning o'z o'rnida qo'llash va qo'llanishini ta'minlashdir. So'zlarni semantik maydonga yig'ib o'rganish ularning turli xususiyati, leksik ma'nosining emotsional ta'siri, boshqa jihatini kuzatib, leksik ma'nosini o'zaro qiyoslab o'rganishda muhim omil sifatida xizmat qiladi [Abdurakhmonova, N. 2021,2022]. Semantik maydonning aniq chegarasi mavjud emas, bu tabiiy hol, chunki bir so'z turli pozitsiyada turlicha talqin qilinishi mumkin. Masalan, **olma mevalar nomi**

semantik maydoni va **olma** – *daraxtlar nomi* semantik maydoni yoki *o'simliklar dunyosi* mavzuviy qurshovi tarkibida hamishtirok etishi mumkin.

Har bir tilning leksikasi o'sha tilda gapiruvchi uchun faol, nafaol holatda bo'ladi. Faol qatlamga kundalik so'zlashuv so'zlari kiradi, nutq jarayonida u so'zlarni miyamizdan qidirib o'tirishga hojat yo'q, bu so'zni aytishga ehtiyoj tug'ilgan zahotiyoq talaffuz etamiz. Nafaol so'z esa biz biladigan, lekin kundalik nutqimizda kam ishlatadigan so'zlarimizdir. Faol so'zda ma'no, talaffuz orasidagi munosabat ikki tomonlama: M.T. Nafaol so'z orasidagi aloqa bir tomonlama, chunki uni o'qiymiz yoki biror kishi nutqidan eshitamiz, uning nutqimizda faollashuvi tez sodir bo'lmaydi, uni eslash kerak [Abdurakhmonova, N. 2021,2022]. Demak, T→M. Bundan kelib chiqadiki, eshitganimiz, o'qiganimizni tushunishga yordam beruvchi lug'at (alifbo lug'ati) foydali, so'zni butun mohiyati, sistemasi bilan ko'z oldimizga keltiruvchi lug'at (ideografik lug'at) esa ma'lum bir til imkoniyatidan bemalol, oson foydalanish imkoniyatini beradi. Shuning uchun bunday lug'at ham muhim axborot banki sirasiga kiradi. Yu.N. Karaulov ideografik lug'atga shunday ta'rif beradi: "Barcha tashkil qiluvchi birliklari orasida kuchli semantik munosabat yaqqol ko'rinuvchi lug'atni tezaurus deb ataymiz" [Карaulов Ю.1990]. Oddiy izohli lug'atdan biror so'zni qidirmoqchi bo'lsak, uning moddiy tomoni(harflar birikmasi)ga e'tibor beramiz. Bunday lug'atga kalit bitta: alifbo. Yu.N.Karaulov alifbo lug'ati, tezaurus haqida shunday misol keltiradi: "D.N.Ushakov lug'atida **dissoniruyushiy** so'zi *dissonirovat, dissotsiatsiya* so'zlari orasida keladi. Bu so'z musiqa sohasiga oid, bu sohaga tegishli barcha so'zlar haqida bilimga ega bo'lish uchun butun lug'atni o'qib chiqish lozim. X.Kasares tezaurusi195da esa bu so'z "Musiqa" maqolasi ichida quyidagi sifatlar qatorida keladi: "... мелодичный, гомофонический, консонирующий, созвучный, аккордийный, гармонический, диатонический, полухроматический, энгармонический, полифалический, характерный, фугированный...". Bu maqola 508 so'zni qamrab olgan. Qisqacha aytganda, butun lug'atni emas, undagi atigi 1-2 betni varaqlasak, kerakli so'z topiladi. Demak, tezaurusda sistema, semantika markazda turadi. Kalit so'zlar esa alifbo tartibida bo'ladi. Ideografik lug'atning yana bir ahamiyatli tomoni shundaki, u turli tillar leksikasini qiyosiy o'rganishda muhim ahamiyatga ega, tarjima lug'ati tuzishda qo'l keladi. Chunki ma'lum tildagi narsa bir so'z bilan atalsa, boshqa tilda bir necha so'z bilan izohlanuvchi birlikka to'g'ri kelishi mumkin. Yevropada ideografik lug'atga qiziqish uyg'otgan "Ingliz so'z, iboralari tezaurusi" ("Тезаурус английских слов и выражений") muallifi P.M.Rodjet o'z lug'ati so'z boshisida aytadi: "Bog'li matn tuzayotgan yozuvchi yoki xohlagan kishi nima demoqchiligini yaxshi bila turib, ayni paytda o'rinli ibora yoki so'zni topolmasa, boy, kerakli so'z tez topiladigan, to'g'ri sistemalashtirilgan, guruhlariga bo'lingan bir kitobdan yordam oladi. Bir so'z bilan aytganda, tezaurusga murojaat etadi... Biz aytmoqchi bo'lgan fikrimizni qanchalik aniq tasavvur qilmaylik, ba'zan shunday bir vaziyatga tushib qolamizki, bizga lug'at boyligimizdagi so'z yetishmay qoladi. O'z fikrimizni aniq ifodalash uchun bitta so'z o'rniga bir necha ijobiy, salbiy, kuchli, kuchsiz, ba'zan esa keraksiz so'zlarni ishlatib yuboramiz. Ideografik lug'atning eng muhim xususiyati shundaki, u foydalanuvchiga bir fikrning barcha qirralarini ochib beruvchi so'z, iboralarning boy to'lasini tortiq qiladi". Demak, ideografik lug'at maqolalari odatdagidek alifbo tartibida emas, balki ma'no (bosh so'z, so'z birikmaning leksik ma'nosi) asosida tartib berilgan lug'atdir. Alifbo tartibli lug'atdan aynan bir so'z ma'nosini bilish mumkin bo'lsa, ideografik lug'at biror tushuncha haqida, bir ma'noni qanday so'zlar bilan ifodalash mumkinligi borasida ma'lumot beradi[Abdurakhmonova, N. 2021,2022].

Leksikografiya ideografik lugʻatlar tuzish borasida katta yutuqlarga erishdi, uning ideografik tezaurus, analogik lugʻat, oʻquvtematik lugʻat kabi turlarini yaratdi. Leksikografiya tarixida ingliz, fransuz, nemis, ispan tillari materiali asosida tuzilgan lugʻatlar anchagina. Bu siraga rus

leksikograflari K.Plots, K.Babov, A.Vorgulev, L.Rozkovsova, Z.Ganusova, V.V.Morkovkin, Yu.N.Karaulov, nemis leksikografi G.O.Fogt, ingliz leksikograflari Dj.Roudel, P.M.Rodjetlar ishlarini kiritish mumkin. Ideografik lugʻatning ahamiyati korpusda ideografik razmetkaga ehtiyoj borligini koʻrsatadi. Muallif asari mavzularini tasnif qilish natijasida birliklar orasidagi mantiqiy aloqadorlik asosida birlashgan soʻzlar toʻplamini aks ettiruvchi ideografik lugʻat paydo boʻladi. A.P.Chexov nasriy, dramatik asarlarining mavzuviy razmetkasi xususiyatlari E.V.Surovseva tomonidan alohida tadqiq etilgan. Bunday mavzuviy razmetkani amalga oshirish uchun uning 583ta nasriy, 17 ta dramatik asari matni tanlangan. Tahlilga olingan matnlar uchun muallifning mukammal asarlari toʻplami (Чехов А.П. Полное собрание сочинений и писем. В 30 томах. М.: Наука, 1974-1983.) asos qilib olingan. Ideografik teglash uchun avtomatik algoritmi – maxsus dastur boʻlmaganligi sababli teglash qoʻlda amalga oshiriladi. Chunki har bir muallif asari mavzu doirasi turlicha boʻladi. Avval muallifning barcha asarlari matni tahlil qilinib, keyin umumiy xulosaga kelinadi. Mavzuviy teg matn mavzusi asosida leksik-semantik guruhlar shakllantirishni maqsad qilib belgilaydi. Matn tanlanib, undan maʼlum bir mavzu atrofida birlashgan soʻzlar belgilanadi. Soʻzlarni belgilashda matnni qayta aytib berish uchun butun matn mazmunini toʻliq ifodalashga xizmat qiluvchi soʻz, iboralar ajratiladi. Aynan shu matndan boshqa bir leksik-semantik guruhga (LSG) mansub soʻzlar alohida rang bilan belgilanadi. Masalan, E.V.Surovseva A.P.Chexov asarlari ideografik razmetkasi uchun yozuvchining “Semiz va oriqlar” hikoyasini obyekt sifatida tanlaydi. Birinchi bosqichda hikoyadan *uchrashuv*, ikkinchi bosqichda *bolalik*, *xotira* mazmuniy guruhiga kiruvchi soʻzlarni alohida ranglar bilan belgilaydi, *doʻstlik* mazmuniy guruhiga kiruvchi soʻzlarni kursivda ajratadi. Shunday tahlildan keyin koʻrinadiki, LSG bir xil emas. Bu guruhlar *holat*, *mavhum tushuncha*, *qahramon tipi* yoki *voqea joyi* kabi guruhlar boʻlishi mumkin. Demak, bitta hikoyada bir necha LSGga kiruvchi soʻzlar guruhi aniqlandi. Faqat bitta LSGga kiruvchi soʻzlardan iborat asar (hikoya) kam uchraydi. Masalan, “Bahorni kutish”da faqatgina *uchrashuv*, “Odamning qisqacha anatomiyasi”da *ilm*, “Parcha”da *provinsiya* kabi LSGlar mavjud. Baʼzi hikoyalarda esa oʻnlab LSG aniqlangan. Asarda hikoyachi, qahramonda xayol surish, fikr yuritish, tahlil qilishga moyillik boʻlsa, oʻsha asarni mavzuviy razmetkalash oson kechadi. Barcha asarlar tahlil qilingandan soʻng mavzular ikkiga: makro, mikromavzuga ajratilib, baʼzi mavzular bir vaqtning oʻzida ikki makromavzuga tegishli boʻlib qolishi mumkin. Deylik, *hayit* LSGi mazmuniga koʻra ham *din*, ham *vaqt* makromavzusiga kiritilishi mumkin. E.V. Surovsevaning fikricha, korpusda bu tipdagi razmetkaning mavjudligi rus tili (umuman olganda, barcha tilni)ni xorijiy til sifatida oʻqitishning samaradorligini oshirishda, korpus materialidan taʼlimiy maqsadda foydalanishda ahamiyatli. Chunki mavzulashtirilgan matn xorijiy til taʼlimi jarayonida oʻquv materiali tayyorlashda axborot manbai vazifasini bajaradi. Hatto bu usulda razmetkalangan korpus yordamida jahon mualliflik kartinasini kuzatish qulayligi ham paydo boʻladi. Ideografik razmetkali korpus yordamida bir muallif asarlarini yaxlit holda gʻoyaviy-badiiy tahlil qilishda, u yoki bu mavzuni qaysi usul va vosita bilan ifodalaganligini oʻrganishda ham asqotadi. Masalan, A. Qahhorning inson, maʼnaviyat mavzusi koʻtarilgan barcha asarlarini bir prizmadan oʻtkazish orqali u yoki bu til ifoda vositalarini qoʻllash mahoratini ham oʻrganish

mumkin. Shuningdek, stilistik, lingvo-poetik, kognitiv, pragmatik, lingvomadaniy, konseptologik, diskursiv nuqtai nazardan o'rganishda qo'l keladi.

Xulosa o'rnida shuni aytish mumkinki, har qanday yuksalayotgan fan o'z amaliyotida kompyuterlar taraqqiyotining yutuqlaridan foydalanadi, lekin ushbu fanlar (korpus lingvistikasi, kompyuter lingvistikasi, kompyuter matematikasi, kompyuter fizikasi) kompyutershunoslikning bir qismi sifatida qabul qilinishi joiz emas. Mazkur fanlar tahlil ob'yektini tadqiq etishda kompyuterdan vosita sifatida foydalanadi, xolos. Shuni aytish mumkinki, korpus lingvistikasi kompyuter lingvistikasining alohida bir qismi yoki bo'limi emas. Ular yo'nalish hamda tadqiqot ob'yekti, maqsad va muammolari turlicha bo'lgan mustaqil soha hisoblanadi. Har bir tilning leksikasi o'sha tilda gapiruvchi uchun faol, nafaol holatda bo'ladi [Abdurakhmonova, N. 2021,2022]. Faol qatlamga kundalik so'zlashuv so'zlari kiradi, nutq jarayonida u so'zlarni miyamizdan qidirib o'tirishga hojat yo'q, bu so'zni aytishga ehtiyoj tug'ilgan zahotiyoyq talaffuz etamiz. Nafaol so'z esa biz biladigan, lekin kundalik nutqimizda kam ishlatadigan so'zlarimizdir. Faol so'zda ma'no, talaffuz orasidagi munosabat ikki tomonlama: $M \leftrightarrow T$. Nafaol so'z orasidagi aloqa bir tomonlama, chunki uni o'qiymiz yoki biror kishi nutqidan eshitamiz, uning nutqimizda faollashuvi tez sodir bo'lmaydi, uni eslash kerak. Demak, $T \rightarrow M$. Bundan kelib chiqadiki, eshitganimiz, o'qiganimizni tushunishga yordam beruvchi lug'at (alifbo lug'ati) foydali, so'zni butun mohiyati, sistemasi bilan ko'z oldimizga keltiruvchi lug'at (ideografik lug'at) esa ma'lum bir til imkoniyatidan bemalol, oson foydalanish imkoniyatini beradi. Shuning uchun bunday lug'at ham muhim axborot banki sirasiga kiradi. Yu.N. Karaulov ideografik lug'atga shunday ta'rif beradi [Караулов Ю.1990]: "Barcha tashkil qiluvchi birliklari orasida kuchli semantik munosabat yaqqol ko'rinuvchi lug'atni tezaurus deb ataymiz". Oddiy izohli lug'atdan biror so'zni qidirmoqchi bo'lsak, uning moddiy tomoni(harflar birikmasi)ga e'tibor beramiz. Kompyuter texnologiyalarining rivoji natijasida zamonaviy tilshunoslikda korpus va kompyuter lingvistikasi terminlarini semantik tahlil, korpus lingvistikasi, kritik diskurs tahlil, shuningdek, deskriptiv, qiyosiy va komponent tahlil metodlari asosida tadqiq etish tobora dolzarb ahamiyat kasb etib bormoqda. Tilshunoslik nuqtai nazaridan yondashsak, kompyuter lingvistikasining paydo bo'lishi bugungi kunda fanning bir qancha muammolariga yechim topish imkoniyatini beradi.

Foydalanilgan adabiyotlar:

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ,2013. – 161 с.
2. Идеографический словарь русского языка. (Сост. О.С.Баранов.) – М., 1995. – 4166 статей.
3. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. – М.: ФЛИНТА: Наука, 2013. –128 с.
4. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. – М.: Прометей, 1990.
5. Структурно-функциональная компьютерная модель татарских морфем. – Казань: ФЭН, 2003. – 220 с.
6. Абдурахмонова, Н., & Абдувахобов, Г. (2021). O 'quv lug 'atini tuzishning nazariy metodologik asoslari. *Международный журнал искусство слова*, 4(6).

7. Abdurakhmonova, N. (2021). Formal-Functional Models of The Uzbek Electron Corpus. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, 10(8), 59-66.
8. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
9. Abdurakhmonova, N., Shakirovich, I. A., & O'G'Li, K. N. S. (2022). Morphological analyzer (morfoAnalyse) Python package for Turkic language. *Science and Education*, 3(9), 146-156.
10. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. – М.: ФЛИНТА: Наука, 2013. –128 с.
11. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
12. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011
13. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.
14. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистикасослари. Монография. – Германия: GlobeEdit, 2020. – 250 б
15. Шаров С.А. Представительный корпусрусского языка в контексте мирового опыта // <https://lamb.viniti.ru>
16. www.ruscorpora.ru
17. www.uzbekcorpus.uz
18. www.philol.msu.ru/~lex/corpus
19. <http://www.philol.msu.ru/~lex/corpus>
20. Ataboev N.B. Chomsky's Critics to Corpus and the Problem of Expert Speakers' Linguistic Intuition // *Филология масалалари*. – Тошкент. (4)/2015.– Б. 108-111.
21. Атабоев Н.Б. Корпус тилшунослиги шаклланишининг диахроник тадқиқи // *Бердақ номидаги Қорақалпоқ давлат университетининг Ахборот-номаси*. – Нукус. №4/ 2018. – Б. 147-150.