IQRO JURNALI , IQRO JOURNAL, ИКРО ЖУРНАЛИ
ISSN:2181-4341| wordlyknowledge.uz/index.php/iqro
Volume 24. Issue 01. Oktabr, 2026

O‘ZBEK TILI SEMANTIK ANALIZATORINI YARATISHNING LINGVISTIK ASOSLARI: JAHON, RUS VA O‘ZBEK TILSHUNOSLIGI TAJRIBASI

Nurboyeva Maftuna Vahobjon qizi
Filologiya bo‘yicha falsafa doktori (PhD),
Millat umidi universiteti o‘qituvchisi
https://doi.org/10.5281/zenodo.23164698

Annotatsiya. Maqolada jahon, rus va o‘zbek tilshunosligida korpus lingvistikasining shakllanishi va rivojlanishi, ushbu yo‘nalishga hissa qo‘shgan olimlarning ilmiy qarashlari hamda o‘zbek tili semantik analizatorini yaratishning lingvistik asoslari yoritiladi. Shuningdek, frazeologik birliklarning semantik xususiyatlarini, xususan, omonimiya va polisemiya hodisalarini lingvistik modellashtirishning ahamiyati ochib beriladi.
Kalit so‘zlar: semantik analizator, korpus lingvistikasi, kompyuter lingvistikasi, tabiiy tilni qayta ishlash, lingvistik modellashtirish, frazeologik omonimiya, frazeologik polisemiya, semantik tahlil.

Abstract. The article discusses the formation and development of corpus linguistics in world, Russian, and Uzbek linguistics, the scholarly views of researchers who contributed to this field, and the linguistic foundations for developing a semantic analyzer of the Uzbek language. It also highlights the importance of linguistically modeling the semantic features of phraseological units, particularly the phenomena of phraseological homonymy and polysemy.
Keywords: semantic analyzer, corpus linguistics, computational linguistics, natural language processing, linguistic modeling, phraseological homonymy, phraseological polysemy, semantic analysis.

Аннотация. В статье рассматриваются становление и развитие корпусной лингвистики в мировой, русской и узбекской лингвистике, научные взгляды ученых, внесших вклад в данное направление, а также лингвистические основы создания семантического анализатора узбекского языка. Особое внимание уделяется значению лингвистического моделирования семантических особенностей фразеологических единиц, в частности явлений фразеологической омонимии и полисемии.
Ключевые слова: семантический анализатор, корпусная лингвистика, компьютерная лингвистика, обработка естественного языка, лингвистическое моделирование, фразеологическая омонимия, фразеологическая полисемия, семантический анализ.

Bugungi kunda tabiiy tilni qayta ishlash (Natural Language Processing) texnologiyalarining jadal rivojlanishi o‘zbek tilining elektron resurslarini yaratish, matnlarni avtomatik tahlil qilish hamda til birliklarining semantik xususiyatlarini kompyuter vositasida modellashtirish masalasini dolzarb vazifalardan biriga aylantirmoqda. Xususan, o‘zbek tili semantik analizatorini takomillashtirishda ko‘pma’nolilik, omonimlik, sinonimiya kabi semantik munosabatlarni to‘g‘ri farqlash muhim ahamiyat kasb etadi.

Dunyo kompyuter tilshunosligida semantik analiz va analizator borasida bir qator olimlar ilmiy izlanish olib borishgan. Manbalarda korpus tizimining ilk koʻrinishlari XIII asrga aloqador deya koʻrsatiladi. XVIII asrda esa A.Kruden ismli ilohiyotchi olim “Injil”dagi barcha soʻz va birikmalarni yigʻib, alifbo tartibida joylashtiradi va ularning kitobda qoʻllanilgan oʻrni, bobi, oyatini ham ilova qilib qoʻyadi[1]. Mashhur konkordans namunalaridan yana biri M.Koudon-Klark va S.Iskot tomonidan Shekspir asarlarining muvofiqlik korpusi boʻlib, u ham alifbo tartibidagi kalit soʻz va birikmalardan tarkib topib, har bir kalit soʻzga qaysi asar, qaysi sahna, sahifa, ustunda joylashganligi haqida ma’lumot biriktirilgan.[2]

Jahon va turkiy tilshunoslikda esa korpus tilshunosligiga ilk qadamlar 1970-yillardan boshlab qoʻyilgan boʻlsa, korpus tizimida morfologik-semantik analizator yaratish masalalarida A.V.Dibo, A.K.Asiryan, A.V.Sheymovich, A.R.Gatiatullin, A.M.Bashirov, P.V.Jeltov, N.A.Israilova, P.S.Bakasova, M.G.Malkovskiy, A.S.Starostin, D.N.Younger, Y.A.Kanevskiy, N.V. Kolpakova, N.A.Israilova, P.S.Bakasova, S.Bird, E.Klein, E.Loper, M.Korobov, A.D.Moskvina, D.Orlova, P.V.Panicheva, O.A.Mitrofanova, A.Gareyshina, Y.Grishina, V.Dyachkov, M.Ionov, A.Koroleva, M.Kudrinskiy, A.Lityagina, Y.Luchina, Y.Sidorova, S.Toldova, S.Savchuk, S.Koval[3–11] kabi bir qator olimlar tadqiqotida batafsil yoritilgan.

Jahon tilshunosligida frazeologik birliklar va ularning leksik-semantik turlarining korpus tizimiga tatbiq etilishi boʻyicha toʻlaqonli tadqiqotlar juda kam, faqat frazeologik sathning muayyan turlari, idiomalarni avtomatik tahlil qilish boʻyicha amalga oshirilgan ishlarni uchratish mumkin. Jumladan, R.Simpson, D.Mendis, ingliz tilidagi idiomalar va individual frazemalar paydo boʻlishi, ularni Michigan akadem ingliz tili nutqiy korpus tizimiga tatbiq etish hamda bu orqali iboralarni nutq konstruksiyalarida toʻgʻri qoʻllashni oʻrgatish, natijada kontekst mazmuni boʻyicha xabardorligini oshirish yuzasidan tadqiqotlar olib borishgan[12]. Shu qatorida, S.O.Ebeling, H.Hasselgard kabi olimlar ta’limiy korpus tizimiga frazeologik birliklar hamda ularning semantik imkoniyatlarini tatbiq etish xususida izlanishlar olib borgan boʻlsa[13], H.Gyllstad, B.Volter, S.Granger ham frazeologik birliklarni korpus tizimiga kiritish va shu orqali oʻzga tilli qatlamga ingliz tilidagi idiomalarni yaxlit shaklda taqdim etish boʻyicha tadqiqotlar e’lon qilgan[14–15].

Rus tilshunosligi va rus tili milliy korpus tizimiga diqqat qaratilsa, frazeologizm sohasiga oid bir qator izlanishlarga duch kelish mumkin. D.O.Dobrovolskiy frazeologik birliklarni oʻrganishda parallel korpusdan foydalanish orqali frazemalarning tillararo tavsiflari yanada batafsil ifoda topishi borasida tajribalar olib borgan[16]. M.B.Budilseva, N.S.Novikova, Y.B.Korotova, A.I.Gorojanov kabi olimlar muayyan leksema ishtirokidagi iboralar tizimini shakllantirib, ularni jonli muloqotdagi chastotasiga koʻra klassifikatsiya qilish, natijada korpus tizimiga biror leksik birlik kiritilganda u ishtirok etgan iboralar ham qoʻllanilish koʻlamiga koʻra izoh sifatida berilishi boʻyicha tadqiqotlar oʻtkazishgan[17–19].

Oʻzbek lingvistikasida kompyuter tilshunosligi, tilning umumiy leksikografik bazasini yaratish va bu orqali matnga avtomatik ishlov berish kabi masalalarda M. Mahmudov, M.Ayimbetov, S.Karimov, G.Jumanazarova, A.Babanarov, D.Oʻrinboyeva, A.Norov, S.Muhammedov, N.Abdurahmonova, Sh.Hamroyevalarning[20–27] tadqiqotlari mavjud. Sanalgan izlanishlar kompyuter lingvistikasi yoʻnalishidagi bir qator muammolarni tadqiq etganligi bilan ahamiyatli. Shuningdek, Sh.Gulyamovaning[28] ham polisemik va omonimik

leksemalarni korpus lingvistikasida semantik farqlovchi modellarini yaratish borasida qilgan ilmiy ishlari tadqiqotimiz uchun asosiy namuna bo‘lib xizmat qilmoqda.

Frazeologik birliklarning modellashtirilishi va korpus tizimiga kiritilishi xususida esa tadqiqotlar sanoqli, ammo mazkur yo‘nalishning ilk tadqiqotlarini o‘tgan asrning 70-80-yillaridan boshlab kuzatish mumkin. Iboralarni modellashtirish xususida dastlabki qadamlarni Sh.Rahmatullayev tadqiqotlarida uchratamiz. Tadqiqotchi model istilohi o‘rniga formula so‘zini qo‘llab, iborani tashkil qilayotgan so‘zlarning turli shakllarda kelish holatini formulada ifodalaydi[29]. G.X.Begmatova o‘zbek milliy korpusida idiomalar bazasini yaratish mavzusida tadqiqot olib borgan[30]. Shuningdek, Sh.G‘aniyeva tomonidan ham frazeologizmlarning strukturasi hamda ularni shakliy va mazmuniy modellashtirish masalalarini somatik frazemalarning semantik-sintagmatik formulalarini tuzgan holda tadqiq etilgan[31].

Jahon, rus va o‘zbek tilshunosligida shakllangan korpus va kompyuter lingvistikasi tajribasi o‘zbek tili semantik analizatorini yaratish uchun muhim nazariy va amaliy asos bo‘lib xizmat qiladi. Ayniqsa, frazeologik birliklarning omonimik va polisemantik xususiyatlarini lingvistik modellashtirish ularning kontekstual ma’nosini aniqroq belgilash hamda semantik tahlil samaradorligini oshirish imkonini beradi. Bu esa o‘zbek tilining raqamli lingvistik resurslarini takomillashtirish va tabiiy tilni avtomatik qayta ishlash tizimlarini rivojlantirishga xizmat qiladi.

Adabiyotlar ro‘yxati

1. Cruden, A. A Complete Concordance to the Holy Scriptures of Old and New Testament. – London, 1737. – 856 p.
2. Cowden Clarke, M.V. The Complete Concordance to Shakespeare: being a verbal index to all the passages in the dramatic works of the poet. – London, 1847. – 860 p.
3. Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов хакасского и древнетюркского языков / Научное обозрение саяно-алтая рецензируемый научный журнал Номер 2(08) (Периодичность − 2 раза в год. серия: Филология.Выпуск 2. 2014. – С. 9–31
4. Исраилова Н.А., Бакасова П.С. Морфологический анализатор кыргызского языка // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». 2017. – Труды конференции. В 2-х томах. Т 2. – Казань: Издательство Академии наук Республики Татарстан, 327 с.
5. Мальковский М.Г., Старостин А.С. Модель синтаксиса в системе морфосинтаксического анализа «TREETON» //Тр. междунар. Конф. Диалог 2006. – М., 2006. – С. 481–492
6. Younger D.H. Recognition and parsing of context-free languages in time n3. Information and Control 10 (2). – P.189-208
7. Каневский Е.А., Колпакова Н.В. К вопросу построения морфологического анализатора // Компьютерная лингвистика и интеллектуальные технологии. 1999. Т. 2. – С. 98–106
8. S., Klein E., Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. – Beijing, 2009
9. Korobov M. Morphological Analyzer and Generator for Russian and Ukrainian Languages. Analysis of Images, Social Networks and Texts: 4th International Conference, AIST 2015. –

Yekaterinburg, Russia, April 9–11, 2015. Communications in Computer and Information Science, Springer, 2015.

10. Moskvina A., Orlova D., Panicheva P., Mitrofanova O. Development of the Core for Syntactic Parser for Russian based on NLTK libraries // Труды XIX Международной объединенной научной конференции «Интернет и современное общество». – Санкт-Петербург, 2016
11. Ляшевская О.Н., Астафьева И., Бонч-Осмоловская А., Гарейшина А., Гришина Ю., Дьячков В., Ионов М., Королева А., Кудринский М., Литягина А., Лучина Е., Сидорова Е., Толдова С., Савчук С., Коваль С. Оценка методов автоматического анализа текста: морфологические парсеры русского языка // Компьютерная лингвистика и интеллектуальные технологии. 2010. № 9 (16). – С. 318–326
12. Simpson R., Dushyanthi Mendis. “A Corpus-Based Study of Idioms in Academic Speech.” TESOL Quarterly 37, no. 3 (2003). – P. 419–441. https://doi.org/10.2307/3588398 (murojaat sanasi: 18.03.2025).
13. Ebeling S.O., Hasselgård H. Phraseology in learner corpus research. In S. Granger, G. Gilquin, F. Meunier (eds) The Cambridge Handbook of Learner Corpus Research. – Cambridge: Cambridge University Press, 2015. – P. 207–230.
14. Gyllstad H., Wolter B. Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? Language Learning 66(2), 2016. – P. 296–323.
15. Granger, S. Phraseology, corpora and L2 research. In Granger, S. (ed.) Perspectives on the L2 Phrasicon: The View from Learner Corpora. – Bristol: Multilingual Matters, 2021. – P. 3–21. https://www.multilingual-matters.com/page/detail/Perspectives-on-the-L2-Phrasicon/?k=9781788924856,(murojaat sanasi: 16.03.2025).
16. Добровольский Д.О. Корпусный подход к исследованию фразеологии: новые результаты по данным параллельных корпусов. Вестник Санкт-Петербургского университета. Язык и литература. Том 17, Nomer 3, 2020. – С. 398–411.
17. Budiltseva M.B., Novikova N.S. Жизнь идиомы – определение актуального корпуса фразеологизмов: опыт экспериментального исследования. RUDN Journal of Language Studies, Semiotics and Semantics, 2023, 14(3). – С. 931–945.
18. Кротова Е. Б. Фразеологизмы-конструкции в корпусе: особенности поиска и анализа выдачи (на материа ле немецкого языка) // Лингвистика и методика преподавания иностранных языков. 2022. № 2 (17). https://doi.org/10.37892/2218-1393-2022-17-2-76-90, (murojaat sanasi: 04.02.20205).
19. Горожанов А. И. Алгоритмы поиска фразеологизмов в лингвистическом корпусе с морфологической разметкой. Филологические науки. Вопросы теории и практики. 2024. Том 17. Выпуск 1. – С. 132–138.
20. Махмудов М.А., Пиотровская А.А., Садыков Т. Система машинного анализа и синтеза тюркской словоформы // Переработка текста методами инженерной лингвистики. – Минск, 1982. – С. 121–130
21. Айымбетов М.К. Опыт лингвостатистического анализа лексики и морфологии каракалпакского публицистического текста: Автореф. дисс… канд. фил. наук. – Ташкент, 1987. – 56 с.

22. Муҳаммедов С.А. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дисс… канд. фил. наук. – Ташкент, 1980. – 51 с.
23. Бабанаров А. Разработка принципов построения словарного обеспечения турецко-русского машинного перевода: Автореф. дисс. … канд. фил. наук. – Л., 1981. – 54 с.
24. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – Б. 136
25. Уринбаева Д.Б. Ўзбек фольклори матнларининг лингвостатистик тадқиқи. – Тошкент: Фан, 2010. – Б. 121
26. Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида): Филол. фан. б. фалс. док. (PhD)... дисс. автореф. – Тошкент, 2018. – 49 б.
27. Ҳамроева Ш.М. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. б. фалс. док. (PhD)... дисс. автореф. – Бухоро, 2017. – 49 б.
28. Гулямова Ш. Ўзбек тили семантик анализаторининг лингвистик асослари. Филол. фан. д-ри дисс. – Фарғона, 2022. – 281 б.
29. Раҳматуллаев Ш. Ҳозирги адабий ўзбек тили. Иккинчи китоб. – Тошкент: Mumtoz so‘z, 2010. – 365 б.
30. Бегматова Г. Х. Ўзбек миллий корпусида идиомалар базасини яратиш муаммолари. НамДУ илмий ахборотномаси. 2019-йил 7-сон. – Б. 215–219.
31. Ганиева Ш.А. Ўзбек фразеологизмлари структураси (шаклий ва мазмуний моделлаштириш). Филол. фан. б. фалс. док. (PhD)... дисс. автореф. – Фарғона, 2017. – 50 б.