O'ZBEK TILI MILLIY KORPUSI TADQIQOTLAR MAVZUSI SIFATIDA

Rabbimova Gulchehra Toyir qizi
Samarqand davlat universiteti
O'zbek tili, jurnalistika va
kompyuter lingvistikasi kafedrasi
stajyor o'qituvchisi
Tel: +9989(33)2969923
gulchehraqoshmonova@gmail.com

Received: 30.10.2025
Revised: 31.10.2025
Accepted: 01.11.2025

Kalit so'zlar:
Kompyuter, korpus lingvistikasi, korpus, lug'at, milliy korpus, lingvistik ta'minot, tabiiy tilni qayta ishlash.

Keywords:
Computer, corpus linguistics, corpus, dictionary, national corpus, linguistic support, natural language processing.

Ключевые слова:
Компьютер, корпусная лингвистика, корпус, словарь, национальный корпус, лингвистическое обеспечение, обработка естественного языка.

Annotatsiya. Mazkur maqolada so'nggi yillarda o'zbek tilshunosligida yil sayin rivojlanib borayotgan yangi yo'nalish – korpus lingvistikasining ahamiyati, tavsifi, tadqiq darajasi, o'rganilishi, olib borilayotgan tadqiqotlar haqida umumiy fikr-mulohazalar beriladi.

Annotation. This article presents general reflections on the significance, description, level of research, and study of a new direction that has been rapidly developing in recent years in Uzbek linguistics — corpus linguistics. It also discusses the ongoing research in this field.

Аннотация. В данной статье представлены общие размышления о значении, характеристике, степени изученности и исследовании нового направления, которое в последние годы стремительно развивается в узбекском языкознании — корпусной лингвистики. Также рассматриваются проводимые исследования в этой области.

O'zbek tilshunosligida kompyuter va korpus lingvistikasi yosh soha, bir necha tadqiqotlar amalga oshirilgan, monografiya, o'quv qo'llanma, darslik va lug'atlar nashr qilingan. Yurtimizning turli oliygohlarida o'z faoliyatlarini olib borgan va borayotgan

ko‘plab tilshunos olimlarimiz amaliy tilshunoslikning ushbu yo‘nalishlarida qator ilmiz izlanishlar olib bordilar. Ana shunday olim va olimalarimiz qatoriga hech shubhasiz 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‘minov, S.Umirova, M.Tursunov, B.Xolmuhammedov kabilarni kiritishimiz mumkin.

Keyingi yillarda Toshkent davlat o‘zbek tili va adabiyoti universitetida faoliyat yuritayotgan tilshunos olimlar hamkorlikda V.Zaxarov, B.Mengliyev, Sh.Hamroyevalar sohada izlanish olib borayotgan tadqiqotchilar va talabalar uchun o‘quv qo‘llanma ishlab chiqdilar. “Korpus lingvistikasi” nomli ushbu o‘quv qo‘llanma Toshkentda 2023-yilda nashr etilib ilm ahliga e’lon qilindi. Kitob korpus lingvistikasi nazariyasi, tashkil topishi, taraqqiyot yo‘li, va bugungi holati bilan tanishtiradi; korpus texnologiyalarini o‘rganish, korpuslar bilan ishlash ko‘nikmasini paydo qilish imkonini yaratadi. Mazkur adabiyot 185 sahifadan iborat bo‘lib, uch bobdan iborat: korpus lingvistikasiga kirish (bu bobda fanning maqsad va vazifalari, predmet va obyekti haqida, korpusning amaliy ahamiyati, undagi asosiy tushunchalar, til korpuslarining turlari va tipologiyasi haqida so‘z boradi), korpusdan foydalanish (mazkur bobda esa korpus lingvistikasi yo‘nalishining bugungi holati: jahon va turkiy tilshunoslikda, parallel korpuslar, mualliflik korpuslari haqida ma’lumotlar beriladi), korpus tuzish texnologiyasi (so‘nggi bobda korpus tuzish texnologiyalari ya’ni bosqichlar, razmetka texnologiyalari, tabiiy tilni qayta ishlash, lingvistik va ekstralingvistik razmetka, korpusning qidiruv tizimi ekanligi haqida axborot almashiniladi). Shu o‘rinda eslatib o‘tish joizki, korpus – elektron yig’ilgan matnlar bazasidir. Korpus tarkibiga kiradigan matnlar quyidagi turlarda bo‘lishi mumkin:

1) alohida bir muallif yoki bir necha yozuvchi asaridan olingan matnlar;
2) ma’lum davr (bir necha o‘n yillik yoki yuz yillik)ni qamrab olgan matnlar;
3) belgilangan mavzudagi zamonaviy matnlar;
4) til va jamiyatning bugungi holatini aks ettiruvchi matnlar. [1.20-b].

Shu bilan birgalikda, Sh.Hamroyevaning yana bir kitobi e’tiborga molik. “Korpus lingvistikasi atamalarining qisqacha izohli lug’ati”. Ushbu lug’at terminologik lug’at hisoblanib, ijtimoiy-gumanitar va axborot texnologiyalari sohasi talabalari, tadqiqotchilar va sohaga qiziquvchilar ommasiga mo‘ljallangan. Lug’atda korpus lingvistikasining eng asosiy terminlari, jumladan, korpus va uning turlari, xususan, mualliflik korpuslari, o‘zbek tili mualliflik korpuslari yaratishning lingvistik asoslari bilan bog’liq atamalar, jahon korpus lingvistikasi terminlari, masalan, lemma, teg, razmetka, annotatsiya, korpus menejeri kabilarning ta’rifi, shuningdek, internet tarmog’idagi mavjud bo‘lgan va ishlayotgan ma’lum korpuslar haqida ma’lumotlar

sharhi aks ettirilgan. Kitobning mundarijasi ikki qismdan iborat bo‘lib, ular til korpusi va uning amaliy ahamiyati hamda foydalanilgan adabiyotlar ro‘yxatidan tashkil topgan. Ushbu ilmiy adabiyotda atamalar izohi tor va keng holatda tushuntirib berilgan. Misol uchun, milliy korpusga quyidagicha ta’rif berilgan: Milliy korpus –janridan qat’iy nazar tilning milliy boyliklari jamlangan electron qidiruv tizimi. [2. 39-b].

Mazkur yosh sohaning o‘zbek tilshunosligida shakllanib taraqqiy etishida N. Abdurahmonova ham salmoqli hissa qo‘shdi. Olima o‘z tadqiqotlari asosi sifatida kompyuter lingvistikasi va korpus lingvistikasini tadqiq etdi va bir qancha ilmiy ishlarni ilm ahliga, ommaga taqdim qildi. Bular sirasiga quyidagi izlanishlarni sanab o‘tish foydadan holi emas, bizningcha: N.Abdurahmonova Inglizcha matnlarni o‘zbek tioliga tarjima qilish dasturining lingvistik ta’minoti (Sodda gaplar misolida) N.Abdurahmonova “Zamonaviy lingvistik korpuslarning kompyuter modellari” “Foreign Languages in Uzbekistan”, 2020, No 1(30), 50-58 https://journal.fledu.uz ISSN: 2181-8215 (online), Abduraxmonova N. Matnlarni kompyuterda ishlash muaommolari,N.Abduraxmonova. Inglizcha sodda gaplarni o‘zbek tiliga tarjima qilish dasturining lingvistik ta’minoti, N.Abdurahmanova. Matnlarni ovozlashtirish dasturining lingvistik ta’minoti: Magistrlik dissertatsiyasi.

Bundan tashqari, E.Xonnazarov ham o‘z tadqiqotida korpus masalasiga to‘xtaladi. “O‘zbek tili korpusi uchun zamonni ifodalovchi grammatik shakllarni annotatsiyalash” mavzusidagi dissertatsiyasi ham soha uchun muhim nazariy manba bo‘lib xizmat qilmoqda. Tadqiqotning kirish qismida korpus haqida quyidagi fikrlar keltirib o‘tilgan. Dunyo tilshunosligida XX asr oxirlarida til korpuslarini tuzish va ularni lingvistik annotatsiyalash dolzarb masalalardan biriga aylandi. Bu esa har bir tilga oid so‘zlarni, grammatik shakllarni teglashning dasturiy va lingvistik ta’minotini yaratishni kun tartibiga qo‘ydi. Natijada, XXI asrga kelib millionlab so‘zlarni o‘z ichiga olgan til korpuslari paydo bo‘ldi. [3. 5-b].

O‘zbek tili milliy korpusi bo‘yicha yana bir qator olimlar va tadqiqotchilar ilmiy izlanishlarni davom ettirmoqda. Quyida ushbu sohada faoliyat yuritayotgan ba'zi mutaxassislar keltirib o‘tilgan. Baxtiyor Mengliyev – Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti "Amaliy tilshunoslik va lingvodidaktika" kafedrasi mudiri, filologiya fanlari doktori, professor. U o‘zbek tili milliy korpusini yaratish va rivojlantirish bo‘yicha tadqiqotlar olib bormoqda. Botir Elov – Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti "Axborot texnologiyalari" kafedrasi mudiri, texnika fanlari bo‘yicha falsafa doktori. U o‘zbek tili milliy korpusini yaratish jarayonida axborot texnologiyalari bo‘yicha tadqiqotlar olib bormoqda.

Suyun Karimov – Sharof Rashidov nomidagi Samarqand davlat universiteti professori, filologiya fanlari doktori. U o‘zbek tili milliy korpusi modeli, matnlarni avtomatik qayta ishlash, tokenlash, lemmalash va grammatik razmetkalash modellari va algoritmlari bo‘yicha tadqiqotlar olib borgan. Shahlo Hamroyeva – Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti doktoranti, filologiya fanlari bo‘yicha falsafa doktori (PhD). U o‘zbek tili milliy korpusi bo‘yicha tadqiqotlar olib bormoqda. O‘ral Xoliyorov – Termiz davlat universiteti "O‘zbek tilshunosligi" kafedrasi katta o‘qituvchisi. U o‘zbek tili milliy korpusi bo‘yicha tadqiqotlar olib bormoqda. Shuningdek, 2022-yil 31-may kuni Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universitetida "O‘zbek tilining milliy korpusi: muammo va vazifalar" mavzusida xalqaro ilmiy-amaliy konferensiya bo‘lib o‘tgan. Ushbu konferensiyada ham ko‘plab olimlar va tadqiqotchilar o‘zbek tili milliy korpusi bo‘yicha ilmiy ma'ruzalar qilishgan. Ushbu olimlar va tadqiqotchilar o‘zbek tili milliy korpusini yaratish, rivojlantirish va tadqiq etish bo‘yicha muhim ishlarni amalga oshirmoqdalar. Manzura Abjalova – Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti "Axborot texnologiyalari" kafedrasi katta o‘qituvchisi, filologiya fanlari bo‘yicha falsafa doktori. U o‘zbek tili milliy korpusini yaratish va rivojlantirish bo‘yicha tadqiqotlar olib bormoqda. Uning quyidagi ayrim tadqiqotlari soha uchun ham ahamiyatlidir. M.Abjalova. O‘zbek tilidagi matnlarni tahrir va tahlil qiluvchi dasturning lingvistik modullari (rasmiy va ilmiy uslubdagi matnlar tahriri dasturi uchun).

Yana bir izlanuvchi O. Xidirov ham ushbu yosh yo‘nalishning rivoji uchun munosib hissa qo‘shmoqda. Uning “Milliy korpus uchun parsing dasturi yaratishning lingvistik asoslari” nomli PhD dissertatsiyasida ilmiy tadqiqot mazmunini yorituvchi quyidagi fikrlar mavjud: Dunyo tilshunosligida XXI asrga kelib, korpus lingvistikasini o‘rganish harakati jadal tus olmoqda. Kompyuter lingvistikasi yo‘nalishida avtomatik tarjima sifatini yaxshilash, til birliklarini teglash nazariyasi, algoritm hamda lingvistik ta’minotini yaratish, matn tahlil qilish (tagging, parsing spelkker) dasturlari tuzish jahon kompyuter lingvistikasida dolzarb masalaga aylanib bormoqda. Tilshunoslikda, xususan, kompyuter lingvistikasi sohasida korpus birliklarini teglash, teg modellari, teglash algoritmini ishlab chiqish, shu asosda avtomatik sintaktik teglash (parsing) dasturini ishlab chiqishga zarurat sezilmoqda. [4. 5-b]

M.Tursunov. – filologiya fanlari bo‘yicha falsafa doktorligi dissertatsiyasida ham aynan korpus hamda unga aloqador jarayonlar haqida fikr-mulohazalar yuritadi. “O‘zbek lingvistik korpusini yaratish tajribasidan (konkordans, tokenayzer, lemmatayzer, razmetkalash dasturlari asosida)” nomli ilmiy ishida lingvistik korpus tushunchasiga ham alohida to‘xtalib, unga quyidagicha izoh berib o‘tadi. Lingvistik

korpus –bu ma’lum bir tamoyillarga asosan jamlangan va umumiy standartlashtirilgan razmetkaga ega bo‘lgan matnlar majmuidir. Dastlab korpus (“Birinchi darajadagi korpus”) deganda biror bir umumiy xususiyatni (til, janr, muallif, davr) birlashtiadigan matnlar to‘plami tushunilgan. Zamonaviy kompyuter texnologiyalari katta hajmli matnlarga lingvistik ishlov berish jarayonlarini bir necha bor soddalashtirish va tezlatishga imkoniyat yaratdi hamda korpus tushunchasining mazmunini o‘zgartirdi. Zamonaviy korpus – bu muayyan bir tilning electron shakldagi matnlar to‘plamiga asoslangan electron axborot-ma’lumot tizimidir. Har bir korpus matnlar ustida filologik yondashuv bilan ishlashga xizmat qiladigan lingvistik va dasturiy ta’minotga ega. [5. 11-b.]

A.Rahmanova “O‘zbek tili milliy korpusini yaratishda kompyuter usullari” tadqiqot ishida o‘zbek tili milliy korpusiga kiritiladigan matnlar turlari, janrlari klassifikatsiyasi tavsifi ham keltirilgan. Shuningdek, milliy korpusning lingvistik tadqiqotlardagi ahamiyati ta’kidlangan. Reprezentativlik korpus lingvistikasining asosiy tushunchalaridan biri bo‘lib, matnlarning turli davrlarni, janrlarni, mualliflarni qamrab olish xususiyatidir. Tadqiqotlarda korpus lingvistikasi minimum ikki xil tipdagi matn korpusini tahlil qilishini ko‘rsatadi.

Universal korpuslar –nutqiy faoliyatning barcha qirrasini namoyon etadigan yozma va og’zaki matnlar korpusi.

Ijtimoiy nutqning ayrim lingvistik yoki madaniy fenomenlarini o‘z ichiga oladigan korpuslar, masalan, maqollar korpusi, gazeta nutqidagi siyosiy metaforalar.

O‘zbek tilining milliy korpusi universal xarakterga ega bo‘lib, nutqiy faoliyatning barcha ko‘rinishlarini o‘z ichiga oladi.[6. 13-b.]

Yuqoridagilar singari quyidagi mualliflar va ularning hamkorlikdagi jamoasi ham kompyuter va korpus lingvistikasi sohasida samarali faoliyat olib bormoqdalar. Quyida tilshunoslar va ularning ayrim tadqiqotlari haqida ma’lumot beriladi.

Quyida O‘zbek tili milliy korpusi (O‘TMK) bo‘yicha olib borilgan ilmiy tadqiqotlar va maqolalarning ayrimlarini taqdim etaman. Ilmiy maqolalar:

“O‘zbek tilining milliy korpusi” – Xurshida Qodirova. “O‘zbek tilining ta’limiy korpusi va uning imkoniyatlari” – R. Alayev. “O‘zbek tili milliy korpusi uchun N-gram modellining ahamiyati” – D. Maxammadiyeva. “O‘zbek tilining milliy korpusi yaratiladi” – SamDU yangiliklari, “O‘zbek tili milliy korpusida so‘z shaklini leksikografik baza asosida qidiruv imkoniyatlari” – Baxtiyor Mengliyev.

O‘zbek tili milliy korpusining (O‘TMK) tadqiq darajasi hozirgi vaqtda rivojlanish bosqichida bo‘lib, bir nechta muhim yo‘nalishlarda izlanishlar olib borilmoqda. Quyida uning asosiy tadqiq darajalari va holati haqida umumiy ma’lumotlar keltirib o‘tamiz: boshlang‘ich ishlar: O‘zbekiston Fanlar akademiyasi,

O‘zbek tilini rivojlantirish jamg‘armasi, universitetlar (masalan, O‘zMU) tomonidan dastlabki bosqichda O‘zbek tili milliy korpusining loyihasi ishlab chiqilgan.

O‘zbek tili milliy korpusi (O‘TMK) sohasida erishilgan asosiy yutuqlar quyidagilarni o‘z ichiga oladi: rasmiy platformalarning yaratilishi – uzbekcorpus.uz — O‘zbek tilidagi matnlarni izlash, tahlil qilish va tilda qo‘llanilayotgan birliklar ustida tadqiqot olib borish imkonini beruvchi ilk milliy korpus platformasi ishga tushirildi. Matnlar morfologik tahlil bilan belgilangan, bu esa lingvistik izlanishlar uchun qulaylik yaratadi. Ta’limiy korpus yaratilishi – UzSchoolCorpus.uz — O‘zbekiston maktab darsliklariga asoslangan ta’limiy til korpusi yaratildi. Bu korpus tilshunoslik, ona tili va adabiyot fanlarida, til materiallarini chuqur tahlil qilishda katta yordam bermoqda.

Umuman olib qaralganda, qaysi xalq yoki millat o‘zining milliy til korpusini tuzar ekan, o‘sha qatlamda tilning omma orasidagi o‘rni mustahkamlanib boradi. Bu jarayon butun jahon tilshunosligidami yoki tor ma’noda o‘zbek tilshunosligida amalga oshirilishidan qat’iy nazar milliy til korpusini yaratishdan ko‘zlangan maqsad o‘zining aniq va samarali faoliyat ekanligi bilan barcha davrlarda ajralib turadi.

Tuzilgan korpus til tizimini ilmiy o‘rganish – o‘zbek tilining grammatikasi, leksikasi va sintaksisini chuqur tahlil qilish imkonini beradi. Ikkinchidan, sun’iy intellekt va avtomatlashtirilgan tarjima – mashina tarjimasi va matn tahlili kabi texnologiyalarni rivojlantirish uchun zarur vosita bo‘lib xizmat qiladi. Uchinchidan, o‘quv jarayoni va ta’lim tizimi – o‘zbek tilini o‘rganuvchilar uchun sifatli elektron resurslarni yaratishga yordam beradi.

Foydalanilgan adabiyotlar ro‘yxati.

1. Zaxarov V., Mengliyev B., Hamroyeva Sh. Korpus lingvistikasi. – Toshkent,2023.
2. Hamroyeva Sh. Korpus lingvistikasi atamalarining qisqacha izohli lug’ati. – Buxoro,2018.
3. Xonnazarov E. O‘zbek tili korpusi uchun zamonni ifodalovchi grammatik shakllarni annotatsiyalash. Filolog.fan.bo‘yicha falsafa doktori (PhD)…diss. Avtoreferat.. – Toshkent,2023.
4. Xidirov O. Milliy korpus uchun parsing dasturi yaratishning lingvistik asoslari. . Filolog.fan.bo‘yicha falsafa doktori (PhD)…diss. Avtoreferat.. – Jizzax,2021.
5. Tursunov M. O‘zbek lingvistik korpusini yaratish tajribasidan (konkordans, tokenayzer, lemmatayzer, razmetkalash dasturlari asosida). Filolog.fan.bo‘yicha falsafa doktori (PhD)…diss. Avtoreferat.. –Samarqand, 2023.

6. Rahmanova A. O‘zbek tili milliy korpusini yaratishda kompyuter usullari. Filolog.fan.bo‘yicha falsafa doktori (PhD)…diss. Avtoreferat.. – Andijon,2022.