



O‘ZBEKISTON RESPUBLIKASI OLIY TA’LIM, FAN VA  
INNOVATSIYALAR VAZIRLIGI

ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT O‘ZBEK  
TILI VA ADABIYOTI UNIVERSITETI

## “ZAHIRIDDIN MUHAMMAD BOBUR MEROSINING SHARQ DAVLATCHILIGI VA MADANIYATI RIVOJIDA TUTGAN O‘RNI”

mavzusidagi III xalqaro ilmiy-nazariy va  
amaliy anjuman materiallari



2026-yil 12-fevral





**O‘ZBEKISTON RESPUBLIKASI OLIY TA’LIM, FAN VA  
INNOVATSIYALAR VAZIRLIGI**

**ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT O‘ZBEK TILI  
VA ADABIYOTI UNIVERSITETI**

**O‘ZBEKISTON YOZUVCHILAR UYUSHMASI**

**BOBUR NOMLI XALQARO JAMOAT FONDI**

**ZAHIRIDDIN MUHAMMAD BOBUR  
MEROSINING SHARQ DAVLATCHILIGI VA  
MADANIYATI RIVOJIDA TUTGAN O‘RNI**

III xalqaro ilmiy-nazariy va amaliy anjuman materiallari

**Toshkent – 2026**

## KORPUS LINGVISTIKASI: ERISHILGAN NATIJALAR VA ISTIQBOLDAGI REJALAR

### CORPUS LINGUISTICS: ACHIEVED RESULTS AND FUTURE PLANS

**Abjalova Manzura Abdurashetovna,**

Filologiya fanlari doktori (DSc),

Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti  
professori

E-mail: [abjalova.manzura@gmail.com](mailto:abjalova.manzura@gmail.com)

ORCID: 0000-0002-1927-2669

DOI: [10.52773/tsuull.conf.2026./DPNA3932](https://doi.org/10.52773/tsuull.conf.2026./DPNA3932)

**Annotatsiya.** *Til korpuslari ta'lim, lingvodidaktika, tadqiqot olib borish kabi yirik jarayonlarda benazir elektron manba bo'lib qoldi. Uning elektron lug'atlardan afzalligi korpus bazasidagi matn birliklarining grammatik, semantik, uslubiy xususiyatlari bo'yicha tavsiflanishi, leksik birlikning manbalar bo'yicha qo'llanish statistikasining aniqlanishi, hozirda hatto grammatik shakllarning ham unumli, unumsiz holatlarini aniqlash imkoniyati mavjudligi bilan ahamiyatli hisoblanadi. Xo'sh, bu borada o'zbek korpus lingvistikasida qanday nazariy va amaliy tadqiqotlarni bajarildi. Shu xususdagi qisqacha sarhisob mazkur maqolada o'z aksini topdi.*

**Kalit so'zlar:** *o'zbek korpusshunosligi, korpus lingvistikasi, Alisher Navoiy mualliflik korpusi, elektron lug'atlar, NLP, modellar.*

**Annotation.** *Language corpora have become a unique electronic resource in major processes such as education, language didactics, and research. Its advantage over electronic dictionaries is the description of the text units in the corpus base according to their grammatical, semantic, and stylistic features, the determination of the statistics of the use of the lexical unit according to the sources, and the possibility to determine even the productive and ineffective cases of grammatical forms. Well, in this regard, what theoretical and practical studies have been carried out in Uzbek corpus linguistics. A brief summary of this topic is reflected in this article.*

**Keywords:** *Uzbek corpus studies, corpus linguistics, author corpus of Alisher Navoi, electronic dictionaries, NLP, models.*

**Kirish.** Bugungi kun o'quvchisiga yaxshi ayonki, til korpuslari ta'lim jarayonida beqiyos o'quv quroli, ilmiy tadqiqotlarda esa qimmatli elektron obyekt va predmet hisoblanadi. Ma'lumki, korpus tilni o'rganish uchun asos sifatida qo'llaniladigan ma'lum qoidalarga muvofiq tanlangan va ishlov berilgan matnlar to'plamidir. Ular statistik tahlil qilish va farazlarni tekshirish, muayyan tilda tadqiqotlar olib borish uchun asosiy manba bo'lib xizmat qiladi. Milliy korpus tabiiy tilni uning ma'lum bir bosqich yoki bosqichlarida va turli janrlar, uslublar,



ijtimoiy va dialektal holatlarda namoyon etadi. Shunga ko‘ra uning davriy nashrlar korpusi, mualliflik korpusi, poetik korpus, dialektal korpus, sintaktik korpus kabi ko‘rinishlari mavjud [Batafsil: Abjalova, 2022]. Korpus tilshunosligi matn korpuslarini shakllantirish, ularni yaratish tamoyillarini tadqiq etish, ulardan foydalanish bilan shug‘ullanadigan tilshunoslik bo‘limidir. Shu bois jahon ilmida korpuslarni yaratish, korpusdagi matnlarni qayta ishlash ishlari ancha taraqqiy etgan. Jahon ilmida korpus lingvistikasiga 1960-yillardan ham nazariy, ham amaliy jihatdan tamal toshi qo‘yilgan va mazkur soha davriy ravishda rivojlanganib, takomillashgan. Korpus, eng avvalo, so‘zlarning matnda qo‘llanishini ko‘rsatuvchi misollar massivi ekanligi bilan lingvistik tadqiqotlar uchun ahamiyatli hisoblanadi. Korpus yordamida so‘zlarning qaysi semantik kategoriyaga mansubligini aniqlash, har bir kategoriyadagi ma‘nosi bilan tanishish mumkin. Semantik izohga ega korpus lingvistik tadqiqotlarda morfologik va semantik amallar bajarish uchun material to‘plash imkonini beradi. Shu nuqtayi nazardan korpus ilmiy-amaliy ahamiyatga ega dinamik harakatdagi lingvistik majmua hisoblanadi.

Bugungi kunda esa keng profilli imkoniyatlarga ega til korpuslarini taqdim etib kelmoqda. Bu borada o‘zbek tilshunosligida korpus lingvistikasining tarixiy asosi XX asrning 80-yillarida yuzaga kelganligini ta’kidlash o‘rinli. Tadqiqotlarni amalga oshirish uchun lingvistik manba – matnlar majmuasi, aniqroq aytganda, matn korpusi shakllantirilgan. Umuman, lingvistik tadqiqotlar uchun material vazifasini bajargan kartotekalardan yirik matnlar to‘plamiga o‘tilishi va ularning hisoblash texnikasida raqamlashtirilishi natijasida o‘zbek tilshunosligida XX asrning 80-yillaridan o‘zbek matnlari korpusi yuzaga kelgan [Batafsil: Abjalova, 2020: 74-76], undan til va nutq birliklari haqida kvantitativ ma’lumotlar, leksema, so‘z shakllari, grammatik shakllarning chastotasini aniqlash, turli davrlarda chastotalar va kontekstlarning o‘zgarishini kuzatish, leksik birliklarning birgalikda uchrashishi haqida ma’lumot olish uchun foydalanilgan. Zamon zayli bilan til korpuslari takomillashib, ko‘plab qulay imkoniyatli interfeyslarga ega holda yaratila boshlandi. Bugungi kunda o‘zbek tilining zamon talablariga javob beruvchi bir qancha korpuslari yaratilmoqda. Hatto lug‘aviy birliklardan foydalanish va ularning semantik xarakterini tezkor aniqlashda qulay bo‘lsin uchun korpuslar bazasida leksikografik manbalar [Abjalova, 2021: 12-17] maxsus shakllantirilmoqda.

**Asosiy qism.** Shu kunga qadar o‘zbek korpusshunosligida ko‘plab tadqiqotlar amalga oshirildi, amaliy natijalarga erishildi, hozirda takomillashtirish yo‘lidan borilmoqda. Ammo hali oldinda dolzarb masalalar qad rostlab turibdi. Shunday bo‘lsa-da, erishilgan natijalarni sarhisob qilish, kelgusida takroriy tadqiqotlarni qilishdan saqlaydi. Til korpusida matnlarning qayta ishlanishi (NLP vazifasi) juda muhim. Shu bois 2009-yildan buyon olib borilgan tadqiqotimizning 2019-yilda tugal holga keltirilishi [Abjalova, 2020:monografiya] ko‘p muammolarning bartaraf etilishida imkon berdi.

Tadqiqotimiz natijasida o'zbek tilidagi matnlarni tahrir va tahlil qilish dasturi modullari yaratishda tabiiy tilni qayta ishlash (NLP) bosqichlaridan: grafematik, morfologik va sintaktik tahlil bo'yicha quyidagi natijalarga erishilgan:

- O'zbek tilining ta'limiy korpusi (O'TTK) [Abjalova va boshq. 2023]da (Guvohnoma № BGU 00514) **so'z** (tadqiqotning grafematik tahlil bosqichi) [Abjalova, 2018: 151-156], [Abjalova, 2019: 1005-1007] va **gap** (tadqiqotning sintaktik tahlil bosqichida berilgan gap sxemalari asosida) [Abjalova, 2020: 89] **tokenizatsiyasi** amalga oshirilgan [Abjalova, 2020:58-63];

- Lemmatizatsiya va stemmitizatsiya jarayonlari aniqlashtirildi; o'zbek tilidagi lemma va stem farqlab berilgan [Abjalova, 2020:25-27];

- O'TTKda monografik tadqiqotdagi o'zbek tilida mavjud tovush o'zgarishlari (tushish, ortish, almashinish) va istisnoli holatlar bo'yicha tuzilgan algoritmlar asosida morfologik tahlil [Abjalova, 2019:422-423], amalga oshirildi [Abjalova, 2019:422-423], [Abjalova, Iskandarov, 2020:160-163], [Abjalova, Iskandarov, 2021:49-52];

- O'TTKni yaratish jarayonida so'z turkumlari bazasi (37000 leksik birlik) identifikatsiyalangan;

- O'rganilgan leksikografik tizimlar asosida O'TTKdagi “Lug'atlar” rukniga asos solindi. Unda tadqiqotimiz natijasidagi **omonimlar** (Guvohnoma № BGU 00381. 2019), **antonim** (Guvohnoma № BGU 00390. 2020) va **paronimlar** (Guvohnoma № BGU 00469. 2021) bazasi kiritilgan.

- Turkiy tillarning lingvodidaktik elektron platformasi [Normurodova va boshq., 2021:15-20] uchun sinonimlar bazasi (Guvohnoma № BGU 00380) maxsuslashtirildi, sinonimlarning turk va qirg'iz tillaridagi tarjimalari berildi;

- Omonim va ko'p ma'noli so'zlarni aniqlashning bigram va trigram (N-gramm) modellari yaratilgan [Abjalova, 2020:117-128], [Abjalova, 2020:666-668], [Abjalova, 2020:14-16], [Abjalova, Yuldashev, 2021], [Abjalova, 2021:1016-1021].

- Avtomatik morfologik va semantik tahlil uchun o'zbek tilidagi so'z yasovchi va grammatik kategoriyalar tizimlashtirilgan.

- Turkiy xalqlar fenomeni Alisher Navoiy mualliflik korpusi yaratilgan (Guvohnoma № DGU 18544) va uning semantik bazasi shakllantirilgan (Guvohnoma № BGU 00583).

- O'zbek tilining multikorpusi (<http://uzcorpora.uz/>) (Guvohnoma № DGU 56015) yaratilib ishga tushirildi. Mazkur korpus o'z tarkibida 15dan ortiq til korpusi turlari bazasiga ega va unda korpus menejeri ham yaratilgan.

Leksikografik bazalar nafaqat korpus tarkibida, balki mobil ilovalar holatida ham foydalanuvchilarga taqdim etildi. Jumladan, o'zbek tilining paronimlari lug'ati, o'zlashma so'zlar lug'ati, o'zbek tilining sinonimlari lug'atini [Dadaboyev va boshq., 2023] mobil qurilmaga Play market xizmatidan yuklab olib foydalanish mumkin.

**Alisher Navoiy mualliflik korpusini** yaratishdan maqsad Alisher Navoiy asarlari korpusini yaratish natijasida mutafakkir ijodini ta'lim resursi sifatidagi ahamiyatini oshirish, ta'lim jarayonidagi mushtarak tahliliy baza ekanligini dalillash hamda g'azalarda bugungi davr tilida tushunilishi qiyin bo'lgan leksik birliklar ma'nolarining kontekstual semantik tavsiflari bazasini yaratib, Navoiy ijodini o'qishli qilish va o'zbek tilining tarixiy modelini yaratishdan iborat.

Alisher Navoiy asarlarini o'qiydiganlar auditoriyasini kengaytirish, foydalanuvchilarda g'azallarini o'qib tushunish kompetensiyasini rivojlantirish maqsadida ham boshlangan Alisher Navoiy mualliflik korpusiga oid tadqiqot ishi yakunlangan bo'lib [G'ulomova, 2023], korpusning o'zi yaratilgan [<http://alishernavoicorpus.uz/uz>], ayni damda amaliy loyiha qilinib<sup>4</sup>, korpusning bazasi Alisher Navoiyning boshqa asarlari semantik teglari bilan boyitilmoqda.

**O'zbek tilining multikorpusini** (<http://uzcorpora.uz/>) interfeysi 4 til – o'zbek, ingliz, rus, turk tillaridan iborat. 4 xil qidiruv imkoniyati mavjud. Hozirda ushbu korpus bazasida 21 mln.dan ortiq so'zshakl bor. Korpusda 20 turdagi til korpusi turlarining bazalari tayyorlangan. Jumladan, ta'limiy korpus, gazetalar (unda “Ma'rifat” gazetasining 9 yillik material mavjud), internet matnlari korpusi, badiiy matnlar korpusi, poetik matnlar korpusi, ilmiy manbalar korpusi, tarixiy manbalar korpusi, rasmiy matnlar korpusi, mualliflik korpusi, parallel korpus kabi korpus turlari bazalari joylangan. Shu bois ushbu korpus o'zbek tilining multikorpusi, ya'ni ko'pturli bazaga ega yirik platforma hisoblanadi.

Eng asosiysi, O'zR Prezidentining 20.10.2025-yildagi 308-son Qarorida ko'rsatilgan 2 ta vazifa ijrosini ta'minlash maqsadida korpusda “Lug'atlar” rukni yaratildi. Unda 1) 2500 sinonimik qatorga ega sinonimlar bazasi; 2) 2 mingta omonimshaklga ega omonimlar bazasi; 3) 600 ta paronimik juftlik mavjud paronimlar bazasi; 4) o'zbek tiliga 27 tildan o'zlashgan 11 ming so'z mavjud o'zlashmalar bazasi tayyorlandi. Hozirda ushbu ruknning “Terminologik lug'atlar” qismi uchun 20ga yaqin sohaga doir terminlar bazasi tayyorlanmoqda.

uzcorpora.uz – o'zbek tilining ko'p maqsadli til korpuslari mujassamlashgan multikorpus; bu lingvistik va adabiy korpuslar platformasini yaratish va rivojlantirishga qaratilgan loyiha bo'lib, u tilshunoslik tadqiqotlari, ta'lim jarayoni va tabiiy tilni qayta ishlash (NLP) sohasida keng foydalaniladi.

Xulosa tarzida aytish joizki, ilmiy maktabimizda amaliy tilshunoslik, korpusshunoslik, tabiiy tilni qayta ishlash va lingvistik ontologiyalarni yaratish masalasida tadqiqot ishlari qilinmoqda. Yaqin vaqtlarda yana yangiliklar bilan o'zbek korpus lingvistikasi boyitilishiga hissa qo'shish har bir mutaxassis burchi hisoblanadi.

### **Adabiyotlar:**

1. Abjalova M. Korpus lingvistikasi. Darslik. – Toshkent: Nodirabegim, 2026. – 350 b.

<sup>4</sup> Foydalanish yo'riqmasi: <https://youtu.be/XCXV-cgJl5M?si=mSyYeW665H5u9Z7b>.

2. Abjalova M. Korpus lingvistikasi: uslubiy qo‘llanma / M.A. Abjalova. – Toshkent: Nodirabegim, 2022. – 110 b.
3. Abjalova M. Lingvistik dasturlar uchun omonimlarni tahlil qilish texnologiyasi // “Tilshunoslikdagi zamonaviy yo‘nalishlar: muammolar va yechimlar” mavzusidagi Xalqaro ilmiy-amaliy 1-onlayn konferensiya. – Andijon, 2020. – B. 666-668.
4. Abjalova M. Matnni grafematik tahrirlash modulining lingvistik normalari// Ilm sarchashmalari. – Urganch, 2018. – №3. – B. 151-156.
5. Abjalova M. Milliy korpusi mavjud bo‘lmagan tillarning lingvistik dasturlarida omonimlarni tahlil qilish texnologiyasi. // So‘z san‘ati. Xalqaro jurnal. – Samarqand, 1/2020. – B. 117-128. ISSN 2181-9297 DOI <http://dx.doi.org/10.26739/2181-9297>;
6. Abjalova M. O‘zbek Milliy korpusining tarixiy asosi. “Filologiyaning dolzarb muammolari // The actual problems of Philology” mavzusidagi Xalqaro ilmiy-amaliy internet konferensiya. – Farg‘ona, 2020. –B. 74-76.
7. Abjalova M. O‘zbek tili korpusining dasturiy ta‘minoti imkoniyatlari. // Kompyuter lingvistikasi: muammolar, yechim va istiqbollar. Xalqaro ilmiy-amaliy konferensiya to‘plami. Elektron nashr / ebook. – Toshkent: ToshDO‘TAU, 22.04.2022. – B. 73-81.
8. Abjalova M. Omonimiya va lingvistik tizimlarda omonimlarni farqlash usullari. // “Oriental Renaissance: Innovative, Educational, Natural And Social Sciences (ORIENS). Vol.1, Issue 10. SJIF 2021-5.423. – pp. 1016-1021. [www.oriens.uz](http://www.oriens.uz) ISSN 2181-1784.
9. Abjalova M. O‘zbek tili milliy korpusida so‘zshaklni leksikografik baza asosida qidiruv imkoniyatlari // Kompyuter lingvistikasi: muammolar, yechim, istiqbollar / Respublika ilmiy-texnik konferensiya to‘plami. Elektron nashr / ebook. – Toshkent: ToshDO‘TAU, 2021. – B. 12-17.
10. Abjalova M. The linguistic norms of the grapheme editing module //International Journal of Science and Research (IJSR), India. ISSN: 2319-7064. Volume 8 Issue 2, 2019. – P. 1005-1007. (№ 40; Research Gate: 0.28) [www.ijsr.net](http://www.ijsr.net).
11. Abjalova M. Принципы создания морфологического модуля для программы автоматического анализа узбекского языка // “Science and education in the modern world: Challenges of the XXI century” – Nur-Sultan, Kazakhstan, July 2019. – R. 422-423.
12. Abjalova M., E. Adalı and O. Iskandarov, "Educational Corpus of the Uzbek Language and its Opportunities," 2023 8th International Conference on Computer Science and Engineering (UBMK), Burdur, Turkiye, 2023, pp. 590-594, doi: 10.1109/UBMK59864.2023.10286682.
13. Abjalova M., Elova D. Tabiiy tilni qayta ishlash (NLP)da so‘z turkumlarini teglash masalasi. // O‘zbekistonda til va madaniyat, – Toshkent: ToshDo‘TAU, 1/2021. – B. 6-20.

14. Abjalova M., Iskandarov O. Lingvistik dasturning morfologik tahlil moduli (morfoanalizator). // FAN va JAMIYAT Ilmiy-uslubiy jurnal, – Nukus, 2/2021. – B. 49-52.
15. Abjalova M., Iskandarov O. Matnlarni morfologik tahlil qilish moduli va uning algoritmi. – T.: O‘zMU xabarlarini // Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti ilmiy jurnali. – Toshkent, 2/2020. – B. 160-163.
16. Abjalova M., Yuldashev A. Methods for Determining Homonyms In Homonymy And Linguistic Systems. ACADEMICIA: An International Multidisciplinary Research Journal. Vol. 11, Issue 2, February 2021. Impact Factor: SJIF 2021 = 7.492 (<https://saarj.com>). ISSN: 2249-7137.
17. Abjalova M.A. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – 176 b. ISBN 978-9943-6939-0-6.
18. “Alisher Navoiy mualliflik korpusini yaratish” nomli amaliy loyiha. Bajirilishi 2023-2024-yillarga mo‘ljallangan.
19. Dadaboyev H., Abjalova M., Rashidova U. O‘zbek tili sinonimlarining o‘quv lug‘ati. – Toshkent: BookPrint, 2023. – 810 b.
20. G‘ulomova N. Alisher Navoiy mualliflik korpusi va uning semantik teglari bazasini yaratish (“Badoye’ ul-vasat” devoni asosida): Fil.f.f.d. (PhD) diss. – Toshkent, 2023. – 160 b.
21. <https://youtu.be/744o55Tqx9E>
22. Normuradova N., Raupova L., Abjalova M. “Turkiy tillarning lingvodidaktik elektron platformasini yaratish” nomli amaliy loyihalar xususida. // “O‘zbek tili taraqqiyoti va xalqaro hamkorlik masalalari” mavzusidagi xalqaro konferensiyasi “Zamonaviy leksikografiya, til korpuslari va turkiy tillar platformalarini yaratish muammolari” nomdagi sho‘ba materiallari, 2021-yil 18-oktabr. – B. 15-20.
23. Абжалова М. Технология анализа омонимов для программ анализа текстов на узбекском языке // Международная научно-практическая конференция “Наука и образование в современном мире: вызовы XXI века”. – Нур-Султан, Казахстан: Молодой ученый, – 2020. 13-17 марта 2020 г. – С. 14-16.
24. <http://alishernavoicorpus.uz/uz> – Alisher Navoiy mualliflik korpusi
25. <http://uzcorpora.uz/> – o‘zbek tilining multikorpusi