



Shohida Xoshimovna
Shahabitdinova
filologiya fanlari doktori, professor
E-mail: shohidas3@mail.ru

TILSHUNOSLIK TADQIQOTLARIDA KORPUS TEXNOLOGIYALARI VA SUN‘IY INTELEKT: RAQAMLI FILOLOGIK TAHLIL MEZONLARI

May

2026 Annotatsiya Mazkur maqolada tilshunoslik tadqiqotlarida axborot texnologiyalari, korpus metodologiyasi, tabiiy tilni qayta ishlash, elektron lug‘atchilik, avtomatik tarjima, nutqni tanish va sun‘iy intellekt asosidagi matn tahlilining ilmiy qiymati yoritiladi. Tadqiqotda 180 ta ilmiy-uslubiy manba va raqamli platforma tavsifidan tuzilgan 42 600 tokenli tahliliy korpus asosida raqamli vositalarning filologik izlanishlardagi qo‘llanish ko‘lami baholandi. Natijalarga ko‘ra, korpus qidiruvi 24.7%, morfologik tahlil 17.3%, semantik teglash 14.9%, avtomatik tarjima 13.5%, avtoreferat chiqarish 9.6%, nutq/OCR vositalari 8.2% va savol-javob tizimlari 5.4% ulush bilan qayd etildi.
Kalit so‘zlar: kompyuter lingvistikasi, korpus metodologiyasi, tabiiy tilni qayta ishlash, elektron lug‘at, sun‘iy intellekt, avtomatik tarjima, nutq texnologiyalari, lingvistik belgilash.

May

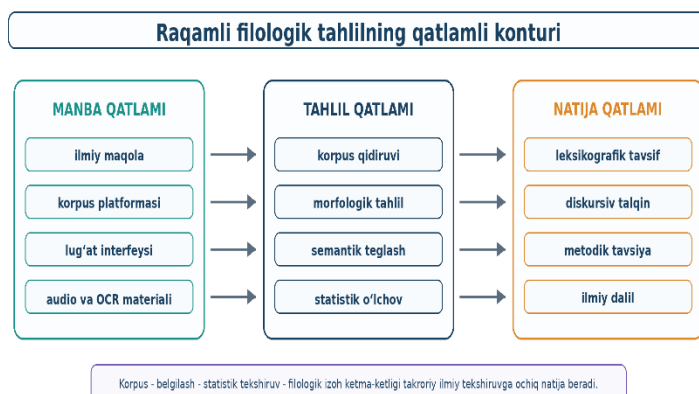
2026 Abstract The article examines information technologies, corpus methodology, natural language processing, electronic lexicography, machine translation, speech recognition, and AI-based text analysis in contemporary linguistic research. The study relies on an analytical corpus of 42,600 tokens compiled from 180 scholarly-methodological sources and digital platform descriptions. The findings show that corpus search accounts for 24.7%, morphological parsing for 17.3%, semantic tagging for 14.9%, machine translation for 13.5%, automatic summarizing for 9.6%, speech/OCR instruments for 8.2%, and question-answering systems for 5.4%.
Keywords: computational linguistics, corpus methodology, natural language processing, electronic dictionary, artificial intelligence, machine translation, speech technologies, linguistic annotation.

KIRISH. Raqamli filologiyabugungi tilshunoslikning eng sermahsul yo‘nalishlaridan biri sifatida shakllanmoqda, chunki til materialini endi faqat kuzatuvchi xotirasi yoki alohida matn namunalari orqali emas, katta hajmli elektron korpuslar, avtomatik belgilash tizimlari, statistik o‘lchovlar va sun‘iy intellektga tayangan tahlil muhitlari orqali ham tekshirilmoqda. Bu holat lingvistik xulosaning daliliy bazasini kengaytirib, matn, nutq, semantika va diskurs qatlamlarini bir vaqtda qiyoslash imkonini beradi.

Kompyuter lingvistikasi til materialini formallashtirish, algoritmik tavsiflash va avtomatik qayta ishlashga yo‘naltirilgan fanlararo soha sifatida maydonga chiqdi. Uning tadqiqot predmeti til bo‘lsa-da, tekshiruv usuli matematik-statistik o‘lchov, dasturiy belgilash, korpusdan qidiruv, avtomatik tasnif va leksikografik tavsifdan iborat murakkab ilmiy tartibni qamrab oladi. Korpus metodologiyasi haqiqiy matnlar asosida farazni sinashga sharoit yaratgani sababli, u zamonaviy tilshunoslikda introspektiv mulohazani empirik tekshiruv bilan to‘ldiradi [1, P.314].

Matn korpuslari filologik izlanishlarda tilning kichraytirilgan, tartiblangan va qayta tekshiriladigan modeli sifatida baholanadi. Korpus yordamida so‘z shakllari chastotasi, kollokatsion birikmalar, grammatik qurilmalar, terminologik o‘zgarishlar, tarixiy siljishlar va diskurs registrlari aniq ko‘rsatkichlar orqali tekshiriladi. Pertsov korpus ma’lumotlari grammatik va leksik da’volarni sinashda odatiy ilmiy amaliyotga aylanishi lozimligini ta’kidlagan [2, P.320]; Fillmore esa kichik korpus ham tadqiqotchiga kutilmagan dalil bera olishini ko‘rsatgan [3, P.40].

Maqolaning maqsadi tilshunoslikka oid tadqiqotlarda axborot texnologiyalarining ilmiy qo‘llanish maydonini aniqlash, korpus vositalari va sun‘iy intellekt yondashuvlarining qaysi tadqiqot amallarida faol ko‘rinishini ko‘rsatish hamda raqamli filologiyauchun metodik tavsiyalar ishlab chiqishdan iborat. Ushbu maqsad asosida uch savol qo‘yildi: birinchidan, korpus texnologiyalari qaysi lingvistik vazifalarda ustuvor ko‘rinadi; ikkinchidan, sun‘iy intellekt vositalari matn va nutq tahliliga qanday ilmiy qo‘shimcha beradi; uchinchidan, raqamli vositalarni qo‘llashda semantik aniqlik va leksikografik izoh qanday nazorat qilinishi kerak.



1-rasm. Raqamli filologik tahlil konturi.

TADQIQOT METODOLOGIYASI. Tadqiqot materiali sifatida 2020-2024-yillar oraliq‘ida e‘lon qilingan ilmiy maqolalar, korpus platformalari bo‘yicha tavsiflar, elektron lug‘at

interfeyslari, tabiiy tilni qayta ishlash bo'yicha o'quv manbalari, avtomatik tarjima va nutq texnologiyalariga oid loyiha izohlari tanlandi. Yakuniy tahliliy korpus 180 matn birligi va 42 600 tokenidan tashkil topdi. Materiallar mavzu, janr, texnologik yo'nalish va lisoniy vazifa bo'yicha kodlandi.

Tahlil uch bosqichda olib borildi. Birinchi bosqichda matnlar korpus qidiruvi, morfologik tahlil, semantik teglash, avtomatik tarjima, avtoreferat chiqarish, nutq/OCR vositalari va savol-javob tizimlari kabi yo'nalishlarga ajratildi. Ikkinchi bosqichda har bir yo'nalish bo'yicha chastota, ulush, qo'llanish sohasi va filologik natija baholandi. Uchinchi bosqichda raqamli vositalar yordamida olinadigan natijalarning leksikografik tavsif, diskursiv talqin va o'quv jarayoniga tatbiq etilish imkonlari tahlil qilindi.

Korpusda uchraydigan termin va texnologik birliklarni baholash uchun Raqamli filologik qo'llanish ko'rsatkichi - RFQK joriy qilindi. RFQK to'rt mezonidan iborat: 1) real matn bilan ishlash darajasi; 2) lingvistik belgilash imkoniyati; 3) statistik qayta tekshirish qulayligi; 4) lug'aviy yoki metodik natijaga aylantirish salohiyati. Har bir mezon 0-25 ball bilan baholandi, umumiy natija 100 ballik shkala orqali aniqlandi. Bunday tartib korpus tilshunosligida takroriy tekshiruv va o'lchov barqarorligi muhimligini ta'kidlovchi yondashuvlarga mos keladi [4, P.28; 5, P.53].

1-jadval. Tahliliy korpus tarkibi va manba qatlamlari.

Manba turi	Matnlar soni	Ulush (%)	Token miqdori	Tadqiqot yo'nalishi
Korpus tilshunosligi bo'yicha ilmiy maqolalar	52	28.9	13 900	chastota, kollokatsiya, kontekst
NLP va avtomatik tarjima tavsiflari	38	21.1	8 700	matnni qayta ishlash, tarjima bahosi
Elektron lug'at va tezaurus interfeyslari	27	15.0	6 300	lug'aviy izoh, termin maydoni
Nutq, OCR va audiodiskurs vositalari	25	13.9	5 800	ovoza asoslangan lisoniy tahlil
Sun'iy intellektga oid ta'limiy manbalar	22	12.2	4 700	o'quv platformasi, avtomatik qayd
Metodik va nazariy adabiyotlar	16	8.9	3 200	metodologik asos, tahlil mezonlari
Jami	180	100.0	42 600	tahliliy korpus

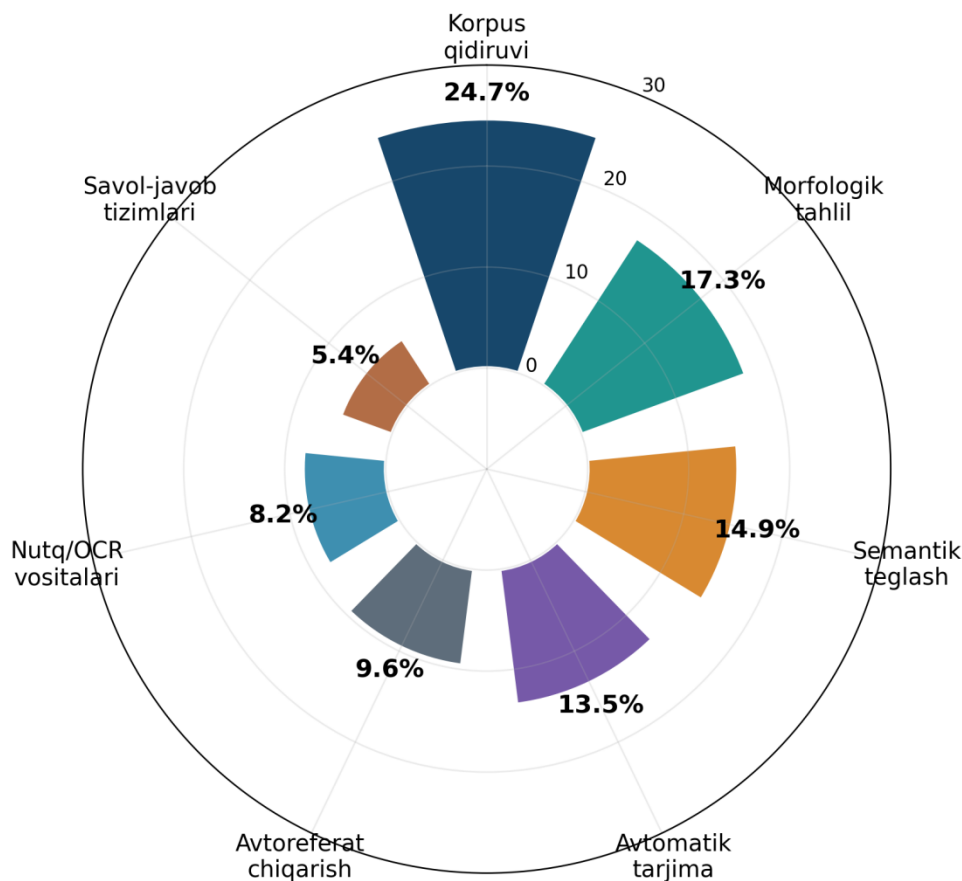
TAHLIL VA NATIJALAR. Tahlil korpusida raqamli filologik amallar teng taqsimlanmaganligi ko‘rindi. Korpus qidiruvi 24.7% ulush bilan yetakchi o‘rinni egalladi, chunki ilmiy matnlarda chastota, kollokatsiya, kontekst va janr bo‘yicha tekshiruv eng ko‘p qayd etilgan. Morfologik tahlil 17.3%, semantik teglash 14.9%, avtomatik tarjima 13.5% natija berdi. Avtoreferat chiqarish, nutqni tanish, OCR va savol-javob tizimlari nisbatan pastroq ulushga ega bo‘lsa-da, ular matnni ixchamlashtirish, ovozli material bilan ishlash va axborot izlashda muhim yo‘nalish sifatida ko‘rindi.

2-jadval. Raqamli yo‘nalishlarning ulushi va filologik natijasi.

Raqamli yo‘nalish	Ulush (%)	RFQK balli	Filologik natija
Korpus qidiruvi	24.7	88	real kontekst va chastota dalili
Morfologik tahlil	17.3	80	so‘z shakli va grammatik qurilma tavsifi
Semantik teglash	14.9	82	ma‘no maydoni va termin chegarasi
Avtomatik tarjima	13.5	69	qiyosiy matn va semantik farqni ko‘rish
Avtoreferat chiqarish	9.6	66	katta matndan qisqa ilmiy qayd olish
Nutq va OCR vositalari	8.2	62	audio hamda rasmi matnni yozuvga ko‘chirish
Savol-javob tizimlari	5.4	58	matndan maqsadli axborot ajratish

Korpusdan olingan sonli ko‘rsatkichlar shuni bildiradiki, tilshunoslik tadqiqotlarida texnologik vositalar bir xil darajada qo‘llanmaydi. Masalan, korpus qidiruvi va morfologik tahlil nazariy tilshunoslik, leksikografiya va grammatik tavsifda ustuvor bo‘lsa, nutqni tanish va sintez qilish fonetika, audiodiskurs va inklyuziv ta‘lim materiallarida ko‘proq uchraydi. Elektron lug‘at va tezauruslar esa terminologik izoh, sinonimik maydon, kollokatsion tavsif hamda so‘z ma‘nosi chegarasini aniqlashda muhim manba bo‘ladi [6, P.47].

Tilshunoslik tadqiqotlarida raqamli vositalar ulushi



2-rasm. Raqamli vositalar ulushining foizli spektri.

RFQK baholovida korpus qidiruvi 88 ball, semantik teglash 82 ball, morfologik tahlil 80 ball, elektron lug‘at va ontologik tavsif 76 ball, avtomatik tarjima 69 ball, nutq/OCR vositalari 62 ball, savol-javob tizimlari 58 ball natija berdi. Ushbu ballar raqamli vosita qancha keng tarqalganini emas, uning filologik tekshiruvda qayta o‘lchanadigan dalil bera olish darajasini ko‘rsatadi. Masalan, avtomatik tarjima keng tarqalgan bo‘lsa-da, semantik noaniqlik va kontekstga bog‘liq tanlovlar sababli u har doim lingvistik xulosa uchun yakka asos bo‘la olmaydi [9, P.211].

Sun‘iy intellektga tayangan tahlil matnlarni mavzuga ajratish, terminlarni topish, gap tuzilishini aniqlash, semantik yaqinlikni o‘lchash va katta hajmli matn oqimini qisqa ilmiy qaydga aylantirishda qulaylik yaratadi. Transformer arxitekturasiga tayangan modellar matndagi uzoq masofali bog‘lanishlarni qayd eta olishi bilan ajraladi [10, P.6002], BERT turidagi modellar esa kontekstga tayangan so‘z ifodasini hisobga olgani uchun ko‘p ma‘noli birliklarni tahlil qilishda alohida ahamiyat kasb etadi [11, P.4173]. Shu bilan birga, bu vositalar tilshunos nazoratini chetga surmasligi, aksincha, dalilni saralash va tekshirish bosqichini kuchaytirishi lozim.

MUHOKAMA. Natijalar axborot texnologiyalarining tilshunoslikdagi o‘rni oddiy yordamchi amaliyot bilan chegaralanmasligini ko‘rsatadi. Korpus, elektron lug‘at, avtomatik tarjima, nutqni tanish va matnning semantik belgilash vositalari til materialini ko‘rish, qiyoslash va izohlash tartibini o‘zgartiradi. An‘anaviy tahlilda tadqiqotchi alohida misolga tayansa, korpus yondashuvida yuzlab yoki minglab kontekst bir vaqtda ko‘riladi; natijada lingvistik xulosa tasodifiy misolga emas, ko‘p martalik dalilga tayanadi.

Til korpuslari ayniqsa leksikografiya uchun katta ilmiy qiymatga ega. So‘zning lug‘aviy ta‘rifi faqat izoh bilan tugamaydi; uning kollokatsion doirasi, uslubiy belgisi, janrga xosligi, tarixiy uchrash chastotasi va qaysi qo‘shimchalar bilan birikishi ham qayd etilishi kerak. Biber, Conrad va Reppen korpus yondashuvi grammatik va leksik tuzilmalarni real qo‘llanish bo‘yicha tadqiq etishini ta‘kidlaydi [6, P.112]. Shu jihatdan elektron lug‘at tuzishda korpus dalili leksikografik tanlovni ancha aniq qiladi.

Sun‘iy intellekt vositalarini qo‘llashda ikki talab muhim: birinchi talab - ma‘lumot manbasining sifati; ikkinchi talab - natijani filologik ekspertiza orqali qayta tekshirish. Agar korpus notekis tuzilgan bo‘lsa, ya‘ni bir janr yoki bitta uslub haddan tashqari ustun kelsa, avtomatik natija ham shu tarafga og‘adi. Shuningdek, mashina tarjimasi yoki avtomatik xulosa chiqarish matnning terminologik qatlamini kamaytirib yuborishi mumkin. Shu sababli raqamli filologiyatexnologik tezlik va lingvistik aniqlik o‘rtasidagi mutanosiblikni talab qiladi.

Amaliy jihatdan mazkur tadqiqot uch yo‘nalish uchun muhim. Birinchi yo‘nalish - ilmiy tadqiqot: korpus vositalari orqali takroriy tekshiruvga ochiq dalil olinadi. Ikkinchi yo‘nalish - ta‘lim: talabalar real matn asosida grammatik, semantik va leksik birliklarni tahlil qilishni o‘rganadi. Uchinchi yo‘nalish - lug‘atshunoslik: elektron lug‘atlarda so‘zning ma‘nosi bilan birga real kontekst, chastota va birikish maydoni ham beriladi. Shunday yondashuv tilshunoslikni aniq dalil, nazariy talqin va amaliy natija uyg‘unligida rivojlantiradi.

XULOSA. Olib borilgan tahlil tilshunoslik tadqiqotlarida axborot texnologiyalari, korpus metodologiyasi va sun‘iy intellekt bir-biridan ajralgan alohida yo‘nalishlar emas, umumiy raqamli filologik tadqiqot tartibining o‘zaro bog‘langan qismlari ekanini ko‘rsatdi. Korpus qidiruvi, morfologik tahlil, semantik tglash, avtomatik tarjima, nutqni tanish, OCR va savol-javob tizimlari til materialini faqat tez qayta ishlash uchungina emas, lingvistik farazni dalil bilan tekshirish uchun ham zarurdir.

Statistik natijalarga ko‘ra, korpus qidiruvi 24.7%, morfologik tahlil 17.3%, semantik tglash 14.9%, avtomatik tarjima 13.5% ulush bilan yetakchi ko‘rsatkichlarni berdi. RFQK baholovi korpus qidiruvi, semantik tglash va morfologik tahlil filologik tekshiruvda eng barqaror natija berishini ko‘rsatdi. Bu natija tilshunoslikda raqamli vositalarni tanlashda faqat ommaviylik emas, dalil sifati, qayta tekshirish imkoniyati va lingvistik izohga yaroqlilik mezonlari ustuvor bo‘lishi zarurligini anglatadi.

Maqolaning ilmiy yangiligi raqamli filologik qo‘llanish ko‘rsatkichi orqali axborot texnologiyalarining lingvistik tadqiqotdagi amaliy qiymatini baholashga urinishidir. Ushbu ko‘rsatkich matn bilan ishlash darajasi, belgilash sifati, statistik qayta tekshirish va leksikografik natijaga aylantirish mezonlarini birlashtirib, kompyuter lingvistikasi bo‘yicha tadqiqotlarni aniqroq rejalashtirishga asos beradi. Kelgusi izlanishlarda o‘zbek tili uchun ochiq morfologik korpus, terminologik tezaurus, dialektal audioarxiv va ko‘p tilli parallel matn bazalarini kengaytirish ustuvor vazifa bo‘lib qoladi.

ADABIYOTLAR:

1. Красовская Ю. Ю. Роль текстовых корпусов в современных лингвистических исследованиях. Лингвистика, лингводидактика, лингвокультурология: актуальные вопросы и перспективы развития. Минск: БГУ, 2024. С. 313-317.
2. Перцов Н. В. О роли корпусов в лингвистических исследованиях. Корпусная лингвистика 2006. Санкт-Петербург: Изд-во С.-Петербур. ун-та; Изд-во РХГА, 2006. С. 318-331.
3. Fillmore, Charles J. "Corpus Linguistics or Computer-Aided Armchair Linguistics." Directions in Corpus Linguistics, edited by Jan Svartvik, Mouton de Gruyter, 2011, pp. 35-60.
4. Захаров В. П., Богданова С. Ю. Корпусная лингвистика: учебник для студентов направления "Лингвистика". 2-е изд. Санкт-Петербург: СПбГУ, 2013. 234 с.
5. McEnery, Tony, and Andrew Hardie. Corpus Linguistics: Method, Theory and Practice. Cambridge University Press, 2012, pp. 1-294.
6. Biber, Douglas, Susan Conrad, and Randi Reppen. Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press, 1998, pp. 1-300.
7. Bird, Steven, Ewan Klein, and Edward Loper. Natural Language Processing with Python. O'Reilly Media, 2009, pp. 1-504.
8. Manning, Christopher D., and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press, 1999, pp. 1-680.
9. Koehn, Philipp. Neural Machine Translation. Cambridge University Press, 2020, pp. 1-408.
10. Vaswani, Ashish, et al. "Attention Is All You Need." Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 5998-6008.
11. Devlin, Jacob, et al. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." Proceedings of NAACL-HLT 2019, Association for Computational Linguistics, 2019, pp. 4171-4186.
12. O'Hagan, Minako. Translation and Technology: Disruptive Entanglement of Human and Machine. Routledge, 2020, pp. 1-190.