TILSHUNOSLIKDA KORPUS LINGVISTIKASINING ILMIY-NAZARIY ASOSLARI: INGLIZ VA O‘ZBEK TILLARI QIYOSIY TAHLILI ASOSIDA

D.M.Jumaboyeva[1]

1. qizi Lavozimi: Osiyo Xalqaro Universiteti, “Ijtimoiy fanlar va filologiya” kafedrasi assistent o‘qituvchisi

Mas'ul muallif: Jumaboyeva Dildora Munis, qizi Lavozimi: Osiyo Xalqaro Universiteti, “Ijtimoiy fanlar va filologiya” kafedrasi assistent o‘qituvchisi.

e-mail: dildorajumaboyeva98@gmail.com

Received: 04 Noyabr 2026
Accepted: 09 Noyabr 2026
Published: 14 Noyabr 2026

Copyright: © 2026 by the author(s). Licensee Buxoro xalqaro universiteti ilmiy axborotnomasi. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY-4.0) license.

Iqtibos uchun: Jumaboyeva, D. M. TILSHUNOSLIKDA KORPUS LINGVISTIKASINING ILMIY-NAZARIY ASOSLARI: INGLIZ VA O‘ZBEK TILLARI QIYOSIY TAHLILI ASOSIDA / D. M. Jumaboyeva // Buxoro xalqaro universiteti ilmiy axborotnomasi. — 2026. — T. 11, № 1. — B. 390–394. — DOI: 10.67780/buxoro-xalqaro-universiteti-ilmiy-axborotnomasi.2026.da512740.

Annotatsiya. Mazkur maqolada zamonaviy tilshunoslikning muhim yo‘nalishlaridan biri bo‘lgan korpus lingvistikasining ilmiy-nazariy asoslari, shakllanish omillari, metodologik tamoyillari va amaliy imkoniyatlari ingliz hamda o‘zbek tillari qiyosiy tahlili asosida tadqiq etiladi. Korpus lingvistikasining empirik tilshunoslik bilan munosabati, til birliklarini tabiiy nutqiy muhitda o‘rganishdagi metodologik ahamiyati, korpuslarni shakllantirish mezonlari va lingvistik annotatsiyalash masalalari tahlil qilinadi. Corpus-based va corpus-driven yondashuvlarining o‘ziga xos xususiyatlari aniqlashtirilib, ularning leksik-semantik, grammatik va qiyosiy tadqiqotlardagi imkoniyatlari asoslanadi. Tadqiqotning ilmiy yangiligi sifatida ingliz va o‘zbek tillari materiallarini kompleks tahlil qilishga yo‘naltirilgan ko‘p bosqichli korpus-leksikografik konseptual model taklif etiladi. Model leksik birlikning lemma, grammatik shakl, ma’no, ta’rif, ekvivalent, kollokatsiya, kontekst va uslubiy belgi kabi parametrlarini o‘zaro bog‘liq tizimda tavsiflashga asoslanadi. Shuningdek, o‘zbek tilining agglutinativ xususiyatlari, morfologik variantlari, yozuv tizimidagi farqlar va kontekstual ekvivalentlik muammolarining korpus yaratish jarayoniga ta’siri yoritiladi. Olingan nazariy umumlashmalar milliy korpus resurslarini takomillashtirish, qiyosiy leksikografiya va raqamli tilshunoslikni rivojlantirish uchun metodologik asos bo‘lib xizmat qilishi mumkin.

Kalit so'zlar: korpus lingvistikasi, til korpusi, empirik tilshunoslik, korpus metodologiyasi, reprezentativlik, morfologik annotatsiya, lemmatizatsiya, kollokatsiya, konkordans, qiyosiy tahlil, leksikografiya, ingliz tili, o‘zbek tili.

Аннотация. В статье исследуются научно-теоретические основы корпусной лингвистики, факторы её становления, методологические принципы и практические возможности на материале сопоставительного анализа английского и узбекского языков. Рассматриваются взаимосвязь корпусной лингвистики с эмпирическим языкознанием, критерии формирования языковых корпусов, особенности лингвистической аннотации и значение корпусных данных для изучения языковых единиц в естественном контексте. Анализируются подходы corpus-based и corpus-driven, а также их применение в лексико-семантических, грамматических и сопоставительных исследованиях. В качестве научной новизны предлагается многоуровневая корпусно-лексикографическая концептуальная модель, объединяющая параметры леммы, грамматической формы, значения, дефиниции, коллокации, контекста и стилистической характеристики. Особое внимание уделяется агглютинативной природе узбекского языка, проблемам морфологической разметки и контекстуальной эквивалентности. Результаты исследования могут служить методологической основой для совершенствования национальных корпусных ресурсов, сопоставительной лексикографии и цифровой лингвистики.

Ключевые слова: корпусная лингвистика, языковой корпус, эмпирическое языкознание, корпусная методология, репрезентативность, морфологическая аннотация, лемматизация, коллокация, конкорданс, сопоставительный анализ, лексикография.

Abstract. This article investigates the theoretical foundations, methodological principles, and practical applications of corpus linguistics through a comparative analysis of English and Uzbek. It examines the relationship between corpus linguistics and empirical approaches to language research, focusing on corpus design, representativeness, linguistic annotation, and the analysis of naturally occurring language data. The study distinguishes between corpus-based and corpus-driven approaches and evaluates their relevance to lexical, semantic, grammatical, and contrastive research. As its principal scholarly contribution, the article proposes a multilayered corpus-lexicographic conceptual model integrating eight interconnected parameters: lemma, grammatical form, meaning, definition, equivalent, collocation, context, and stylistic label. Particular attention is paid to the agglutinative structure of Uzbek, morphological variation, orthographic diversity, and contextual equivalence in English-Uzbek comparative studies. The proposed framework provides a theoretical basis for improving national corpus resources, contrastive

https://doi.org/10.67780/buxoro-xalqaro-universiteti-ilmiy-axborotnomasi.2026.da512740

lexicography, and digital linguistic research.

Keywords: corpus linguistics, language corpus, empirical linguistics, corpus methodology, representativeness, morphological annotation, lemmatization, collocation, concordance, contrastive analysis, lexicography.

KIRISH

Globallashuv, raqamli texnologiyalar va sun'iy intellektning jadal rivojlanishi zamonaviy tilshunoslik oldiga til hodisalarini yangicha metodologik yondashuvlar asosida o'rganish vazifasini qo'ymoqda. Xususan, til birliklarining real kommunikativ muhitdagi qo'llanish xususiyatlarini katta hajmdagi elektron matnlar asosida tadqiq etish zarurati korpus lingvistikasining ilmiy ahamiyatini yanada oshirmoqda. Bugungi kunda tilni faqat nazariy qoidalar va alohida misollar asosida tavsiflash bilan cheklanib qolmasdan, uning tabiiy qo'llanishini empirik dalillar yordamida o'rganish zamonaviy lingvistik tadqiqotlarning muhim talablaridan biriga aylanmoqda.

Korpus lingvistikasi til birliklarini tabiiy nutqiy materiallar asosida, ularning chastotasi, distributiv xususiyatlari, kontekstual muhiti va o'zaro birikuvchanligini hisobga olgan holda tadqiq etuvchi ilmiy-metodologik yo'nalishdir. D. Biber, S. Conrad va R. Reppen korpusga asoslangan tadqiqotlarni real til materiallari yordamida til tuzilishi va qo'llanish xususiyatlarini aniqlashga yo'naltirilgan metodologik yondashuv sifatida izohlaydilar [1]. Bu yondashuv tilshunoslikda nazariy farazlarni empirik ma'lumotlar bilan tekshirish va lingvistik qonuniyatlarni tabiiy nutqiy muhitda aniqlash imkoniyatini yaratadi.

J. Sinclairning korpus lingvistikasiga oid ilmiy qarashlarida so'zlarning ma'nosi va funksional xususiyatlarini ularning kontekstual muhiti hamda kollokatsion munosabatlaridan ajratmagan holda o'rganish zarurligi asoslanadi [2]. S. Gries esa korpus lingvistikasining muhim metodologik tushunchalari sifatida reprezentativlik, muvozanatlilik, chastota, konkordans va kollokatsion tahlilni ajratib ko'rsatadi [3].

Mazkur nazariy qarashlar ingliz va o'zbek tillarini qiyosiy o'rganishda ham alohida ahamiyat kasb etadi. Chunki ushbu tillar grammatik qurilishi, morfologik xususiyatlari, so'z yasalishi va grammatik ma'nolarni ifodalash vositalari jihatidan sezilarli farqlarga ega. Ingliz tilida grammatik munosabatlarning muayyan qismi yordamchi so'zlar va so'z tartibi orqali ifodalansa, o'zbek tilida agglutinativ morfologik vositalar muhim o'rin tutadi. Natijada ikki tildagi leksik birliklarni oddiy lug'aviy muvofiqlik asosida qiyoslash ularning barcha semantik va funksional xususiyatlarini ochib bera olmaydi.

Shu nuqtayi nazardan, korpus lingvistikasining ilmiy-nazariy asoslarini ingliz va o'zbek tillari materiallari asosida o'rganish, mavjud metodologik yondashuvlarni qiyosiy tahlil qilish hamda leksikografik tadqiqotlar uchun kompleks model ishlab chiqish dolzarb ilmiy muammo hisoblanadi.

Tadqiqotning maqsadi va vazifalari

Tadqiqotning asosiy maqsadi korpus lingvistikasining ilmiy-nazariy va metodologik asoslarini aniqlash, ingliz va o'zbek tillarida korpus tahlilining o'ziga xos xususiyatlarini qiyosiy o'rganish hamda korpus ma'lumotlarini leksikografik tavsiflashga yo'naltirilgan konseptual modelni taklif etishdan iborat.

Ushbu maqsadni amalga oshirish uchun quyidagi vazifalar belgilandi:

korpus lingvistikasining shakllanishi va nazariy taraqqiyotini tahlil qilish;

korpus yaratishning fundamental metodologik tamoyillarini aniqlash;

corpus-based va corpus-driven yondashuvlarini qiyoslash;

ingliz va o'zbek tillarida korpus tahlilining lingvistik xususiyatlarini aniqlashtirish;

morfologik annotatsiya, lemmatizatsiya va kontekstual tahlilning ahamiyatini asoslash;

korpus ma'lumotlarini leksikografik tavsiflashning ko'p bosqichli modelini ishlab chiqish;

o'zbek milliy korpus resurslarini takomillashtirish istiqbollarini belgilash.

Tadqiqot obyekti va predmeti

Tadqiqot obyekti sifatida ingliz va o'zbek tillarining elektron korpuslari hamda ularda aks etgan leksik, grammatik va semantik birliklar belgilandi.

Tadqiqot predmetini korpus lingvistikasining nazariy-metodologik tamoyillari, korpus ma'lumotlarini lingvistik qayta ishlash mexanizmlari, qiyosiy tahlil usullari va leksikografik tavsiflash imkoniyatlari tashkil etadi.

ADABIYOTLAR TAHLILI VA TADQIQOT METODOLOGIYASI

Korpus lingvistikasining ilmiy-nazariy asoslari xalqaro tilshunoslikda bir necha bosqichda shakllangan. Dastlabki elektron korpuslardan biri hisoblangan Brown Corpus 1960-yillarda ingliz tilidagi yozma matnlarni kompyuter yordamida tahlil qilish imkoniyatini yaratdi. Keyinchalik korpus materiallarining hajmi, tarkibi va lingvistik annotatsiyalash imkoniyatlari kengayib, korpus lingvistikasi mustaqil tadqiqot metodologiyasi sifatida rivojlandi.

J. Sinclairning Corpus, Concordance, Collocation asari korpus yaratish, konkordans tahlili va kollokatsion munosabatlarni aniqlash masalalarining nazariy asoslarini yoritgan muhim tadqiqotlardan biridir [2]. Muallifning qarashlari so'z ma'nosini uning real qo'llanish muhiti bilan bog'liq holda o'rganish zarurligini ko'rsatadi.

D. Biber, S. Conrad va R. Reppen korpus lingvistikasining grammatik, leksik, diskursiv va funksional-uslubiy tadqiqotlardagi imkoniyatlarini tizimli ravishda asoslaydilar [1]. E. Tognini-Bonelli esa corpus-based va corpus-driven yondashuvlarini farqlab, korpus dalillarining nazariy tilshunoslikdagi o'rnini belgilaydi [4]. T. McEnery va A. Hardie korpus lingvistikasining metod, nazariya va amaliyot bilan bog'liq masalalarini yaxlit ilmiy tizim sifatida tahlil qiladilar [5].

O'zbek tilshunosligida korpus yaratish va parallel matnlarni lingvistik belgilash masalalari ham alohida tadqiqot yo'nalishiga aylangan. B. Elov va M. Amirkulov o'zbek-ingliz parallel korpusini shakllantirish, matnlarni tanlash, lingvistik belgilash va parallel birliklarni moslashtirish bosqichlarini yoritganlar [6]. Shuningdek, o'zbek tilining milliy korpusida token, lemma va konkordans asosida qidirish, morfologik tahlil hamda parallel korpus imkoniyatlari maxsus elektron

platformada namoyon etilgan [7].

Mavjud ilmiy manbalar tahlili korpus lingvistikasining umumiy nazariy tamoyillari yetarlicha ishlab chiqilganini ko‘rsatsa-da, ingliz va o‘zbek tillari o‘rtasidagi morfologik hamda semantik tafovutlarni leksikografik tavsiflash bilan birlashtiruvchi kompleks modelni ishlab chiqish zarurati mavjudligini ko‘rsatadi.

Tadqiqotda tavsifiy, qiyosiy-tipologik, tizimli-struktur, kontekstual, komponent tahlil va nazariy umumlashtirish metodlaridan foydalanildi. Shuningdek, korpus lingvistikasiga oid ilmiy adabiyotlar, mavjud korpus resurslari va ingliz-o‘zbek leksik birliklarining namunaviy qiyosiy tahlili tadqiqotning empirik-metodologik asosini tashkil etdi.

Tadqiqot nazariy-tahliliy xarakterga ega. Unda keltirilgan ingliz va o‘zbek tilidagi leksik birliklar korpusda aniqlangan statistik natijalar sifatida emas, balki taklif etilayotgan metodologik modelning ishlash mexanizmini tushuntiruvchi illustrativ misollar sifatida qo‘llanadi.

NATIJALAR VA MUHOKAMA

1. Korpus lingvistikasining ilmiy-nazariy mohiyati

Korpus lingvistikasini faqat elektron matnlarni to‘plash yoki kompyuter yordamida so‘zlarni qidirish bilan bog‘liq texnik faoliyat sifatida talqin qilish uning ilmiy mohiyatini to‘liq ochib bermaydi. Aslida, korpus lingvistikasi til hodisalarini kuzatish, dalillarni tizimlashtirish, ularni miqdoriy va sifat jihatidan tahlil qilish hamda nazariy umumlashmalar shakllantirishga asoslangan kompleks metodologik yo‘nalishdir.

Korpus lingvistikasining asosiy ilmiy xususiyati tilni uning real qo‘llanishi orqali o‘rganishidir. Bunda til birliklari alohida va ajratilgan holatda emas, balki muayyan matn, kommunikativ vaziyat, janr va nutqiy muhit doirasida tadqiq qilinadi. Mazkur yondashuv tilning tizimli xususiyatlari bilan uning amaliy qo‘llanishi o‘rtasidagi munosabatni aniqlashga yordam beradi.

Korpus lingvistikasining tadqiqot obyekti tabiiy nutqda qayd etilgan til materiallari bo‘lsa, predmeti ushbu materiallarda namoyon bo‘ladigan leksik, grammatik, semantik, sintaktik, uslubiy va pragmatik qonuniyatlardan iborat.

Uning ilmiy-metodologik maqomi empiriklik, tizimlilik, miqdoriylik, kontekstual yondashuv va qayta tekshiriluvchanlik xususiyatlari bilan belgilanadi. Demak, korpus lingvistikasi nazariy tilshunoslik, amaliy tilshunoslik va kompyuter lingvistikasining o‘zaro integratsiyasi natijasida shakllangan ko‘p qirrali ilmiy metodologiya hisoblanadi.

2. Korpus yaratishning fundamental metodologik tamoyillari

Til korpusining ilmiy qiymati uning hajmi bilangina emas, balki tarkibiy jihatdan qanchalik to‘g‘ri shakllantirilgani va tadqiqot maqsadiga mos kelishi bilan belgilanadi.

Birinchi muhim tamoyil – reprezentativlik. Bu korpus tarkibining o‘rganilayotgan til yoki tilning muayyan qatlamini yetarli darajada aks ettirishini anglatadi. Masalan, faqat badiiy matnlardan tashkil topgan korpus asosida ilmiy, rasmiy yoki so‘zlashuv uslubiga xos xususiyatlar haqida umumiy xulosa chiqarish metodologik jihatdan cheklangan bo‘ladi.

Ikkinchi tamoyil – muvozanatlilik. Korpus tarkibida turli janr, uslub, kommunikativ vaziyat va davrlarga oid matnlarning tadqiqot maqsadiga muvofiq nisbatda joylashtirilishi uning muvozanatliligini ta’minlaydi.

Uchinchi tamoyil – autentiklik. Korpusga kiritiladigan matnlar tabiiy kommunikativ jarayonda yaratilgan bo‘lishi lozim. Sun’iy yaratilgan misollar real nutqning barcha xususiyatlarini aks ettira olmaydi.

To‘rtinchi tamoyil – annotatsiyalash. Korpus birliklariga morfologik, grammatik, sintaktik va semantik belgilar biriktirilishi ma’lumotlarni chuqur tahlil qilish imkonini beradi.

Beshinchi tamoyil – metama’lumotlarning mavjudligi. Har bir matnning muallifi, yaratilgan sanasi, janri, manbasi, uslubi va boshqa xususiyatlari qayd etilishi korpusdan foydalanish samaradorligini oshiradi.

Oltinchi tamoyil – qayta tekshiriluvchanlik. Korpus tarkibini shakllantirish mezonlari, qidiruv parametrlari va tahlil usullari aniq belgilanishi tadqiqot natijalarining ilmiy ishonchliligini mustahkamlaydi.

Ushbu tamoyillar bir-birini to‘ldiradi. Reprezentativlik va muvozanatlilik korpusning tarkibiy sifatini belgilasa, annotatsiyalash va metama’lumotlar uning lingvistik tahlil imkoniyatlarini kengaytiradi.

3. Corpus-based va corpus-driven yondashuvlarining nazariy farqlari

Korpus lingvistikasida ikki asosiy metodologik yondashuv mavjud: corpus-based va corpus-driven. E. Tognini-Bonelli ushbu yondashuvlarni nazariy va metodologik jihatdan farqlaydi [4].

Corpus-based yondashuvda tadqiqotchi avvaldan mavjud bo‘lgan nazariy qarash yoki ilmiy farazni korpus materiallari orqali tekshiradi. Masalan, ingliz tilidagi modal fe’llarning turli nutq janrlarida qo‘llanish chastotasini o‘rganish yoki o‘zbek tilidagi kelishik qo‘shimchalarining funksional xususiyatlarini korpus asosida aniqlash ushbu yondashuvga misol bo‘ladi.

Corpus-driven yondashuvda esa tadqiqotning boshlang‘ich nuqtasi tayyor nazariy faraz emas, balki korpusda kuzatilgan lingvistik dalillardir. Tadqiqotchi dastlab birliklarning takrorlanish holatlari, kontekstual munosabatlari va distributiv xususiyatlarini o‘rganadi. Keyinchalik kuzatuvlar asosida nazariy xulosalar shakllantiriladi.

Mazkur yondashuvlarni birgalikda qo‘llash ingliz va o‘zbek tillarini qiyosiy o‘rganishda ayniqsa samaralidir. Chunki nazariy jihatdan ma’lum bo‘lgan grammatik yoki semantik farqlarni korpus ma’lumotlari bilan tekshirish va shu bilan birga, korpusda kuzatilgan yangi qo‘llanish qonuniyatlarini aniqlash mumkin.

4. Ingliz va o‘zbek tillarida korpus tahlilining qiyosiy-metodologik xususiyatlari

Ingliz va o‘zbek tillari o‘rtasidagi tipologik tafovutlar korpuslarni shakllantirish va tahlil qilishda alohida metodologik yondashuvni talab qiladi.

Ingliz tili asosan analitik grammatik xususiyatlarga ega bo‘lib, grammatik munosabatlarning ifodalanishida yordamchi so‘zlar, so‘z tartibi va grammatik konstruksiyalar muhim o‘rin tutadi. O‘zbek tili esa agglutinativ xususiyatga

https://doi.org/10.67780/buxoro-xalqaro-universiteti-ilmiy-axborotnomasi.2026.da512740

ega bo‘lib, grammatik ma’nolar ko‘pincha so‘z negiziga qo‘shiladigan affikslar orqali ifodalanadi.

Masalan, ingliz tilidagi books, book's, books' va o‘zbek tilidagi kitoblar, kitobning, kitoblarimizdan shakllari morfologik jihatdan turlicha tashkil topgan. O‘zbek tilida bitta leksemaning ko‘plab grammatik shakllarda namoyon bo‘lishi korpus qidiruvida lemmatizatsiya va morfologik tahlilning zarurligini ko‘rsatadi.

Shuningdek, ingliz tilidagi make a decision birikmasi o‘zbek tilida qaror qabul qilmoq shaklida ifodalanadi. Ushbu misolda so‘zma-so‘z tarjima emas, balki ikki tilning me’yoriy kollokatsion munosabatlari asosiy ahamiyatga ega. Demak, qiyosiy korpus tahlilida faqat alohida so‘zlarning muvofiqligi emas, balki ularning birikuvchanligi va kontekstual funksiyasi ham hisobga olinishi zarur.

Yana bir muhim jihat – yozuv tizimining farqlanishidir. O‘zbek tilida lotin va kirill yozuvlaridan foydalanish, imlo variantlari va ayrim leksik birliklarning yozilishidagi tafovutlar korpus materiallarini birxillashtirishni talab etadi. Ingliz tilida esa imlo variantlari, qisqartmalar va grammatik shakllarning o‘ziga xos xususiyatlarini hisobga olish zarur.

Shu asosda ingliz va o‘zbek tillari korpuslarini qiyoslashda leksik birlikning boshlang‘ich shakli, morfologik va grammatik variantlari, semantik tarkibi, kollokatsion xususiyatlari, sintaktik vazifasi, kontekstual qo‘llanishi, funksional-uslubiy belgisi va tarjima ekvivalentlarining muvofiqlik darajasi muhim hisoblanadi.

5. Korpus lingvistikasining asosiy tahlil metodlari

Korpus lingvistikasida til materiallarini o‘rganish bir nechta o‘zaro bog‘liq metodlar asosida amalga oshiriladi.

Chastotaviy tahlil ma’lum bir birlikning korpusda necha marta uchrashini aniqlaydi. Chastota ko‘rsatkichlari til birliklarining faolligi, janrlar bo‘yicha taqsimlanishi va qo‘llanish dinamikasini o‘rganishda muhimdir. Biroq korpus hajmi va janr tarkibi hisobga olinmasdan olingan xom chastota ko‘rsatkichlari turli korpuslarni bevosita qiyoslash uchun yetarli emas. Bunday holatlarda nisbiy chastota va normallashtirilgan ko‘rsatkichlardan foydalanish lozim.

Konkordans tahlili muayyan so‘z yoki grammatik shaklning korpusdagi qo‘llanish holatlarini kontekst bilan birgalikda ko‘rsatadi. KWIC (Key Word in Context) usuli yordamida kalit birlikning oldingi va keyingi konteksti ko‘rib chiqiladi. Bu esa ko‘pma’noli so‘zlarning semantik xususiyatlarini aniqlash imkonini beradi.

Kollokatsion tahlil so‘zlarning muntazam birikuvchanligini o‘rganadi. Masalan, strong tea va kuchli choy shakllarini qiyoslash orqali lug‘aviy muvofiqlik bilan tabiiy birikuvchanlik o‘rtasidagi farqni aniqlash mumkin. Ingliz tilidagi make a decision va o‘zbek tilidagi qaror qabul qilmoq birikmalari esa turli leksik komponentlar orqali ifodalangan funksional muvofiqlikni namoyon etadi.

Distributiv tahlil birliklarning turli janrlar, matn turlari va kommunikativ vaziyatlarda taqsimlanishini aniqlaydi. Qiyosiy korpus tahlili esa ikki yoki undan ortiq til korpuslaridagi birliklarning qo‘llanish xususiyatlarini taqqoslashga asoslanadi. Parallel korpuslar yordamida tarjima ekvivalentlari, semantik farqlar va grammatik transformatsiyalar aniqlanadi.

Ushbu metodlarning o‘zaro uyg‘unligi korpus lingvistikasini kompleks tadqiqot metodologiyasiga aylantiradi.

6. Korpus lingvistikasi va zamonaviy leksikografiya integratsiyasi

Korpus lingvistikasining muhim amaliy yo‘nalishlaridan biri zamonaviy leksikografiyadir. An’anaviy lug‘atlarda so‘zlarning ma’nosi, grammatik tavsifi va qo‘llanish namunalari ko‘pincha cheklangan miqdordagi misollar asosida beriladi. Korpusga asoslangan leksikografiyada esa lug‘at maqolasini shakllantirish real nutqiy materiallar, chastotaviy ko‘rsatkichlar, kollokatsiyalar va kontekstual ma’lumotlarga tayanadi.

Bunda leksik birlikning tavsifi bosh so‘z yoki lemma, grammatik va morfologik variantlar, semantik ma’no va ma’no variantlari, lug‘aviy ta’rif, boshqa tildagi funksional ekvivalent, tipik kollokatsiyalar, autentik kontekstual misollar va funksional-uslubiy belgilarni qamrab olishi mumkin.

Masalan, ingliz tilidagi decision leksemasini o‘zbek tilidagi qaror leksemasi bilan qiyoslashda faqat lug‘aviy ekvivalentni ko‘rsatish yetarli emas. Ularning grammatik xususiyatlari, kollokatsion birikmalari, semantik qo‘llanish doirasi va kontekstual farqlarini ham hisobga olish zarur.

Korpusga asoslangan elektron lug‘atlarda bunday ma’lumotlarni o‘zaro bog‘lash foydalanuvchiga so‘zning nafaqat lug‘aviy ma’nosini, balki real kommunikativ muhitdagi funksiyasini ham tushunish imkonini beradi.

7. Mualliflik ilmiy yangiligi: ko‘p bosqichli korpus-leksikografik konseptual model

Tadqiqotning asosiy ilmiy natijasi sifatida ingliz va o‘zbek tillaridagi leksik birliklarni korpus asosida kompleks tavsiflashga yo‘naltirilgan ko‘p bosqichli korpus-leksikografik konseptual model taklif etiladi.

Mazkur model korpus materiallarini leksikografik tavsiflashda bir-biridan ajratilgan ma’lumotlar emas, balki o‘zaro semantik va funksional bog‘liq parametrlar tizimi sifatida ko‘rib chiqishga asoslanadi.

Model sakkizta tarkibiy komponentdan tashkil topadi: lemma – leksik birlikning bosh shakli; grammatik shakl – morfologik va grammatik variantlar; ma’no – semantik tarkib; ta’rif – leksikografik izoh; ekvivalent – tillararo funksional muvofiqlik; kollokatsiya – birikuvchanlik xususiyatlari; kontekst – tabiiy qo‘llanish muhiti; uslubiy belgi – funksional-uslubiy tavsif.

Modelning umumiy ko‘rinishi quyidagicha ifodalanadi: LEKSEMA → LEMMA → GRAMMATIK SHAKL → MA’NO → TA’RIF → EKVIVALENT → KOLLOKATSIYA → KONTEKST → USLUBIY BELGI.

Ushbu ketma-ketlik qat’iy chiziqli jarayonni emas, balki o‘zaro bog‘langan ma’lumotlar tizimini ifodalaydi. Masalan, kontekst tahlili natijasida leksemaning yangi ma’nosi aniqlanishi, kollokatsion ma’lumotlar esa ekvivalentni qayta ko‘rib chiqishni talab qilishi mumkin.

Taklif etilayotgan modelning muhim jihati shundaki, u leksikografik maqolani shakllantirishda faqat so‘zning ma’nosini emas, balki uning grammatik, semantik, kollokatsion, kontekstual va uslubiy xususiyatlarini ham yagona tizimga

https://doi.org/10.67780/buxoro-xalqaro-universiteti-ilmiy-axborotnomasi.2026.da512740

birlashtiradi. Mazkur modeldan ingliz-o‘zbek elektron lug‘atlarini yaratish, parallel korpus materiallarini tahlil qilish, tarjima ekvivalentlarini aniqlash va o‘quv lug‘atlarini takomillashtirishda foydalanish mumkin.

8. O‘zbek milliy korpusini rivojlantirishning istiqbolli yo‘nalishlari

O‘zbek tilshunosligida korpus lingvistikasini rivojlantirish nafaqat nazariy, balki amaliy va texnologik jihatdan ham dolzarb ahamiyatga ega. O‘zbek tilining agglutinativ tuzilishi, ikki yozuv tizimida qo‘llanishi, shevalar xilma-xilligi va zamonaviy kommunikativ muhitning o‘zgarishi milliy korpus resurslarini shakllantirishda alohida metodologik yondashuvni talab qiladi.

Birinchidan, morfologik analizator va lemmatizatsiya tizimlarini takomillashtirish zarur. Chunki bitta leksemaning turli grammatik shakllarda namoyon bo‘lishi qidiruv natijalarining to‘liqligiga bevosita ta’sir qiladi.

Ikkinchidan, lotin va kirill yozuvidagi matnlarni o‘zaro bog‘lash, imlo variantlarini aniqlash va matnlarni standartlashtirish mexanizmlarini rivojlantirish lozim.

Uchinchidan, milliy korpus tarkibida badiiy, ilmiy, publitsistik, rasmiy, og‘zaki va internet diskursiga oid materiallarning muvozanatli taqsimlanishiga e’tibor qaratish zarur.

To‘rtinchidan, parallel korpuslarni rivojlantirish ingliz va o‘zbek tillari o‘rtasidagi leksik-semantik va grammatik munosabatlarni chuqur o‘rganish imkonini beradi. Bunda gaplararo moslik bilan bir qatorda so‘z va morfema darajasidagi mosliklarni ham belgilash maqsadga muvofiq.

Beshinchidan, korpus ma’lumotlarini zamonaviy tabiiy tilni qayta ishlash texnologiyalari bilan integratsiyalash lozim. Bunday integratsiya avtomatik lemmatizatsiya, morfologik tahlil, semantik qidiruv va elektron lug‘atlarni shakllantirish imkoniyatlarini kengaytiradi.

Oltinchidan, korpus materiallaridan foydalanishda mualliflik huquqi, shaxsiy ma’lumotlarni himoya qilish, manbalarni ko‘rsatish va ma’lumotlarning ishonchliligini tekshirish talablariga rioya etilishi zarur.

Shunday qilib, o‘zbek milliy korpusini rivojlantirish lingvistik resurslarni raqamlashtirish bilangina cheklanmay, milliy tilning grammatik, leksik, semantik va uslubiy imkoniyatlarini tizimli ravishda ilmiy hujjatlashtirishni ham nazarda tutadi.

XULOSA. O‘tkazilgan nazariy-metodologik tahlil natijasida korpus lingvistikasi zamonaviy tilshunoslikning empirik dalillar, tizimli kuzatish va kompyuter texnologiyalariga asoslangan muhim ilmiy yo‘nalishi ekanligi aniqlandi. Uning asosiy ilmiy mohiyati til birliklarini tabiiy nutqiy muhitda, real qo‘llanish holatlari va o‘zaro munosabatlari asosida o‘rganishdan iborat.

Tadqiqot natijalari korpus lingvistikasining ilmiy-metodologik asosini reprezentativlik, muvozanatlilik, autentiklik, tizimlilik, annotatsiyalash va qayta tekshiriluvchanlik tamoyillari tashkil etishini ko‘rsatdi. Ushbu tamoyillar korpusning ilmiy sifatini va undan olinadigan natijalarning ishonchliligini belgilaydi.

Corpus-based va corpus-driven yondashuvlari lingvistik tadqiqotlarda bir-birini to‘ldiruvchi metodologik mexanizmlar sifatida namoyon bo‘ladi. Ularning integratsiyasi nazariy farazlarni empirik tekshirish va yangi lingvistik qonuniyatlarni aniqlash imkoniyatini kengaytiradi.Ingliz va o‘zbek tillarining tipologik xususiyatlari korpus yaratish va tahlil qilishda turli metodologik talablarni yuzaga keltiradi. Xususan, o‘zbek tilining agglutinativ tuzilishi morfologik tahlil va lemmatizatsiya tizimlarini takomillashtirishni, ingliz tilining analitik xususiyatlari esa grammatik konstruksiyalar va funksional birliklarni kontekstual o‘rganishni talab qiladi.Chastotaviy, konkordans, kollokatsion, distributiv va qiyosiy tahlil metodlarining o‘zaro uyg‘unligi til birliklarining formal, semantik, funksional va kontekstual xususiyatlarini kompleks o‘rganish imkonini beradi.Tadqiqot doirasida taklif etilgan ko‘p bosqichli korpus-leksikografik konseptual model lemma, grammatik shakl, ma’no, ta’rif, ekvivalent, kollokatsiya, kontekst va uslubiy belgini yagona tizimda birlashtirishga asoslanadi. Ushbu model ingliz-o‘zbek elektron lug‘atlarini yaratish va qiyosiy leksikografik tadqiqotlarni rivojlantirish uchun nazariy asos bo‘lib xizmat qilishi mumkin.O‘zbek milliy korpusini takomillashtirishda morfologik annotatsiya, yozuv variantlarini standartlashtirish, parallel korpuslarni kengaytirish va tabiiy tilni qayta ishlash texnologiyalarini integratsiyalash ustuvor vazifalar sifatida belgilanadi.

Foydalanilgan adabiyotlar ro'yxati

[1] Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating Language Structure and Use. – Cambridge: Cambridge University Press, 1998. – 300 p.

[2] Sinclair J. Corpus, Concordance, Collocation. – Oxford: Oxford University Press, 1991. – 179 p.

[3] Gries S. Th. What is Corpus Linguistics? // Language and Linguistics Compass. – 2009. – Vol. 3, No. 5. – P. 1225–1241. – DOI: 10.1111/j.1749-818X.2009.00149.x.

[4] Tognini-Bonelli E. Corpus Linguistics at Work. – Amsterdam: John Benjamins Publishing Company, 2001. – 224 p.

[5] McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. – Cambridge: Cambridge University Press, 2012.

[6] Elov B., Amirqulov M. Steps of Creating a Tagged Parallel Corpus of the Uzbek-English Languages // Uzbekistan: Language and Culture. – 2022. – Vol. 5, No. 2.

[7] O‘zbek tili milliy korpusi. Elektron resurs. – URL: https://uzbekcorpus.uz/ (murojaat sanasi: 06.11.2026).

[8] Hunston S. Corpora in Applied Linguistics. – Cambridge: Cambridge University Press, 2002.

[9] Kennedy G. An Introduction to Corpus Linguistics. – London: Longman, 1998.

[10] McEnery T., Wilson A. Corpus Linguistics: An Introduction. 2nd ed. – Edinburgh: Edinburgh University Press, 2001.

https://doi.org/10.67780/buxoro-xalqaro-universiteti-ilmiy-axborotnomasi.2026.da512740