QO‘QON DAVLAT PEDAGOGIKA INSTITUTI ILMIY XABARLARI
(2026-yil 10-son)
FILOLOGIYA
PHILOLOGY
https://doi.org/10.70728/b.series.fil.v08.i10.036

LINGVISTIK KORPUSLARDA RAZMETKA TURLARI VA ULARNING TIL HODISALARINI TAHLIL QILISHDAGI AHAMIYATI

Sobirova Zarnigor Raximovna
Filologiya fanlari bo‘yicha falsafa doktori,dotsent,
Buxoro davlat universiteti
https://orcid.org/0000-0001-7550-368X
FBX-0626-0042
z.r.sobirova@buxdu.uz

Annotatsiya. Mazkur maqolada korpus lingvistikasida lingvistik va ekstralingvistik razmetkaning o‘rni, asosiy turlari hamda til hodisalarini avtomatlashtirilgan va statistik tahlil qilishdagi ahamiyati yoritilgan. Matn komponentlarini maxsus teglar yordamida belgilashning nazariy va amaliy jihatlari tahlil qilinib, morfologik, semantik, sintaktik, anaforik, prosodik va diskursiv razmetkaning o‘ziga xos xususiyatlari izohlangan. Shuningdek, mazkur razmetka turlarining leksik, grammatik, semantik va pragmatik hodisalarni tadqiq etish, matnlararo munosabatlarni aniqlash hamda tabiiy tilni qayta ishlash texnologiyalarini rivojlantirishdagi imkoniyatlari asoslangan. Korpus ma’lumotlarining ishonchliligi va tadqiqot natijalarining aniqligi razmetkalash mezonlarining izchilligi bilan bog‘liqligi ko‘rsatib berilgan.

Kalit so‘zlar: korpus lingvistikasi, lingvistik korpus, razmetka, lingvistik teg, ekstralingvistik teg, morfologik razmetka, semantik razmetka, sintaktik razmetka, anaforik razmetka, prosodik razmetka, diskursiv razmetka.

Аннотация. В статье рассматриваются роль лингвистической и экстралингвистической разметки в корпусной лингвистике, её основные виды и значение для автоматизированного и статистического анализа языковых явлений. Анализируются теоретические и практические аспекты маркирования компонентов текста с помощью специальных тегов, раскрываются особенности морфологической, семантической, синтаксической, анафорической, просодической и дискурсивной разметки. Обосновываются возможности применения данных типов разметки при исследовании лексических, грамматических, семантических и прагматических явлений, выявлении межтекстовых связей и разработке технологий обработки естественного языка. Подчёркивается зависимость достоверности корпусных данных и точности исследовательских результатов от последовательности и обоснованности принципов разметки.

Ключевые слова: корпусная лингвистика, лингвистический корпус, разметка, лингвистический тег, экстралингвистический тег, морфологическая разметка,

семантическая разметка, синтаксическая разметка, анафорическая разметка, просодическая разметка, дискурсивная разметка.

Abstract. This article examines the role of linguistic and extralinguistic annotation in corpus linguistics, its principal types, and its significance for the automated and statistical analysis of language phenomena. It discusses the theoretical and practical aspects of annotating textual components with specialized tags and explains the distinctive features of morphological, semantic, syntactic, anaphoric, prosodic, and discourse annotation. The article also explores how these annotation types facilitate the investigation of lexical, grammatical, semantic, and pragmatic phenomena, the identification of textual relations, and the development of natural language processing technologies. Particular attention is paid to the relationship between annotation consistency, corpus data reliability, and the validity of research findings.

Keywords: corpus linguistics, linguistic corpus, annotation, linguistic tag, extralinguistic tag, morphological annotation, semantic annotation, syntactic annotation, anaphoric annotation, prosodic annotation, discourse annotation.

Kirish. Zamonaviy tilshunoslikda axborot-kommunikatsiya texnologiyalarining rivojlanishi til hodisalarini katta hajmdagi elektron matnlar asosida o‘rganish imkoniyatlarini sezilarli darajada kengaytirdi. Korpus lingvistikasi til birliklarining real nutq jarayonida qo‘llanishini kuzatish, ularning chastotasini aniqlash, turli kontekstlardagi funksional xususiyatlarini tahlil qilish va ilmiy farazlarni empirik dalillar asosida tekshirishga xizmat qiluvchi muhim yo‘nalish hisoblanadi.

Lingvistik korpusning ilmiy qiymati faqat uning tarkibidagi matnlar soni yoki hajmi bilan belgilanmaydi. Matnlarning reprezentativligi, tanlash tamoyillari, lingvistik belgilanish darajasi va ma’lumotlarni qayta ishlash usullari ham tadqiqot natijalarining ishonchliligiga bevosita ta’sir ko‘rsatadi. Shu jihatdan korpus tarkibidagi matn komponentlarini maxsus teglar yordamida belgilash, ya’ni razmetkalash, korpus lingvistikasining muhim metodologik bosqichlaridan biridir.

Razmetkalash jarayonida matn birliklariga ularning grammatik, semantik, sintaktik, referensial, prosodik va diskursiv xususiyatlarini ifodalovchi belgilar biriktiriladi. Shuningdek, matnning yaratilgan vaqti, muallifi, janri, manbasi va kommunikativ vaziyati kabi ekstralingvistik ma’lumotlar ham qayd etilishi mumkin. Ushbu ma’lumotlar korpus materiallarini belgilangan mezonlar asosida izlash, saralash, guruhlash va statistik tahlil qilish imkonini beradi.

Tadqiqotning maqsadi lingvistik korpuslarda qo‘llaniladigan razmetka turlarining nazariy xususiyatlarini aniqlash hamda ularning til hodisalarini tahlil qilishdagi ilmiy-amaliy ahamiyatini yoritishdan iborat.

Tadqiqotning asosiy vazifalari lingvistik va ekstralingvistik razmetka tushunchalarini izohlash, morfologik, semantik, sintaktik, anaforik, prosodik va diskursiv razmetkaning funksional xususiyatlarini tavsiflash hamda ularning korpus tadqiqotlaridagi imkoniyatlarini aniqlashdan iborat.

Tadqiqot metodlari. Maqolada tavsifiy, tasniflash, qiyosiy-tahliliy va umumlashtirish metodlaridan foydalanildi. Mazkur metodlar razmetka turlarining vazifalarini aniqlash, ularning o‘zaro munosabatlarini ko‘rsatish va lingvistik tadqiqotlarda qo‘llanish xususiyatlarini yoritishga xizmat qiladi.

Mavzuga oid adabiyotlar tahlili

Korpus lingvistikasining nazariy va metodologik asoslari T. McEnery va A. Hardie, D. Biber, S. Conrad va R. Reppen kabi tadqiqotchilarning ilmiy ishlarida yoritilgan. T. McEnery va A. Hardie korpusga asoslangan tadqiqotlarning nazariy tamoyillari, ma’lumotlarni tahlil qilish usullari va lingvistik xulosalarni empirik dalillar bilan asoslash masalalariga e’tibor qaratadi. D. Biber, S. Conrad va R. Reppen esa turli matn turlarini qiyoslash, grammatik birliklarning qo‘llanish xususiyatlarini aniqlash hamda korpus materiallarini amaliy tahlil qilish usullarini yoritadi.

Korpusdagi lingvistik belgilash masalalari tabiiy tilni avtomatik qayta ishlash va kompyuter lingvistikasi bilan ham uzviy bog‘liq. J. Jurafsky va J. H. Martinning tabiiy tilni qayta ishlashga bag‘ishlangan tadqiqotlarida morfologik va sintaktik tahlil, semantik talqin, referensial munosabatlar hamda diskursni qayta ishlash bilan bog‘liq yondashuvlar yoritilgan.

Lingvistik razmetkaning alohida turlari o‘zaro bog‘liq bo‘lsa-da, ularning har biri muayyan tahlil vazifasini bajaradi. Morfologik razmetka so‘zlarning grammatik xususiyatlarini, sintaktik razmetka esa gap tarkibidagi munosabatlarni aniqlashga xizmat qiladi. Semantik razmetka birliklarning ma’no xususiyatlarini, anaforik razmetka referensial bog‘lanishlarni, prosodik razmetka og‘zaki nutqning ohang va akustik xususiyatlarini, diskursiv razmetka esa matnning kommunikativ tashkil etilishi va birliklar o‘rtasidagi diskursiv munosabatlarni o‘rganishga yordam beradi.

Shu asosda razmetka turlarini alohida-alohida tavsiflash bilan bir qatorda, ularning yagona korpus tizimida o‘zaro uyg‘unlashish imkoniyatlarini ham o‘rganish muhim hisoblanadi.

Asosiy qism

1. Lingvistik va ekstralingvistik razmetkaning mohiyati. Lingvistik razmetka – korpus tarkibidagi matn birliklariga ularning til tizimiga mansub xususiyatlarini ko‘rsatuvchi maxsus teglar yoki belgilarni biriktirish jarayonidir. Ushbu jarayon so‘zlarning grammatik shakllari, semantik xususiyatlari, sintaktik vazifalari, referensial aloqalari, prosodik belgilari va diskursiv funksiyalarini tizimli qayd etish imkonini beradi.

Ekstralingvistik razmetka esa matnning bevosita lingvistik tuzilishidan tashqaridagi, uning yaratilishi va qo‘llanish sharoitini tavsiflovchi ma’lumotlarni qamrab oladi. Jumladan, muallif, yaratilgan sana, matn janri, manba, hudud, kommunikativ vaziyat va boshqa metama’lumotlar ekstralingvistik belgilash tarkibiga kiritilishi mumkin.

Lingvistik va ekstralingvistik razmetka o‘zaro to‘ldiruvchi xususiyatga ega. Masalan, muayyan so‘zning qaysi grammatik shaklda qo‘llanganini morfologik teg orqali aniqlash mumkin. Ushbu shaklning qaysi davrga mansub matnlarda yoki qaysi janrda ko‘proq uchrashini aniqlash uchun esa korpusning tegishli metama’lumotlaridan foydalaniladi.

Razmetkalashning asosiy vazifasi matnni shunchaki raqamli shaklda saqlashdan iborat emas. Uning mohiyati matn tarkibidagi til birliklarini mashina yordamida aniqlash, guruhlash va tahlil qilish mumkin bo‘lgan tizimli ma’lumotlarga aylantirishda namoyon bo‘ladi.

2. Morfologik razmetka. Morfologik razmetka so‘zlarning grammatik xususiyatlarini belgilashga qaratilgan bo‘lib, korpus lingvistikasida eng ko‘p qo‘llaniladigan razmetka turlaridan biridir. Bunda so‘zning turkumi, grammatik shakli, soni, kelishigi, zamoni, mayli, shaxs-son ko‘rsatkichlari va boshqa morfologik belgilar qayd etilishi mumkin.

Masalan, o‘zbek tilidagi “o‘quvchilarimizga” so‘zi quyidagicha tahlil qilinishi mumkin: *o‘quvchi* – ot, *-lar* – ko‘plik, *-imiz* – birinchi shaxs ko‘plikdagi egalik, *-ga* – jo‘nalish kelishigi. Morfologik razmetka yordamida ushbu shakllarning korpusda uchrash chastotasini aniqlash, ularning boshqa grammatik shakllar bilan munosabatini o‘rganish va turli matnlardagi qo‘llanishini qiyoslash mumkin.

Morfologik razmetkaning ahamiyati, ayniqsa, o‘zbek tili kabi qo‘shimchalar orqali ko‘plab grammatik ma’nolar ifodalanadigan tillarda yaqqol namoyon bo‘ladi. Bunday tillarda so‘z shakllarining avtomatik tahlili o‘zak va qo‘shimchalarni farqlash, grammatik kategoriyalarni aniqlash hamda morfologik variantlarni tizimlashtirish uchun muhimdir.

Shu bilan birga, ayrim grammatik shakllarning bir nechta ma’noda qo‘llanishi yoki bir xil shaklning turli grammatik vazifalarni bajarishi morfologik razmetkada kontekstual tahlilni talab qiladi.

3. Semantik razmetka. Semantik razmetka matn tarkibidagi so‘z va birikmalarning ma’no xususiyatlarini belgilashga xizmat qiladi. Bunda birliklarning semantik guruhlari, kontekstual ma’nolari, ma’no munosabatlari va zarur hollarda semantik rollari qayd etilishi mumkin.

Masalan, “ildiz” so‘zi botanika kontekstida o‘simlikning bir qismi, matematika kontekstida esa son yoki ifodaning ildizi ma’nosida qo‘llanishi mumkin. Semantik razmetka ushbu ma’nolarni kontekst asosida farqlash imkonini beradi.

Mazkur razmetka turi ko‘p ma’noli so‘zlarni o‘rganish, sinonim va antonim birliklarni aniqlash, terminologik guruhlarni shakllantirish hamda turli janrlarda ma’no o‘zgarishlarini kuzatishda muhim ahamiyatga ega. Shuningdek, semantik belgilash tabiiy tilni avtomatik qayta ishlash tizimlarida matn mazmunini tahlil qilish va so‘zlarning kontekstga mos ma’nosini aniqlash uchun qo‘llanadi.

Semantik razmetkaning aniqligi teglar tizimining qanday ishlab chiqilganiga, ma’no kategoriyalarining chegaralari qanchalik aniq belgilanganiga va matn kontekstining qay darajada hisobga olinganiga bog‘liq.

4. Sintaktik razmetka. Sintaktik razmetka gap tarkibidagi so‘z va birikmalarning grammatik munosabatlarini hamda gap qurilishini belgilashga qaratiladi. Bunda ega, kesim, to‘ldiruvchi, aniqlovchi va hol kabi sintaktik vazifalar yoki so‘zlar o‘rtasidagi bog‘lanishlar qayd etilishi mumkin.

Masalan, “Talaba ilmiy maqolani diqqat bilan o‘qidi” gapida *talaba* – ega, *o‘qidi* – kesim, *ilmiy maqolani* – to‘ldiruvchi vazifasida, *diqqat bilan* esa harakatning bajarilish tarzini bildiruvchi hol vazifasida keladi.

Sintaktik razmetka gap qurilishining modellarini aniqlash, grammatik konstruksiyalarning qo‘llanish chastotasini o‘rganish, so‘z tartibini qiyoslash va avtomatik sintaktik tahlil tizimlarini ishlab chiqishda muhim vosita hisoblanadi.

Turli tillarning sintaktik xususiyatlarini qiyoslashda ushbu razmetka turi alohida ahamiyat kasb etadi. Chunki tillar so‘z tartibi, grammatik munosabatlarni ifodalash usullari va gap qurilishi jihatidan farqlanishi mumkin. Sintaktik belgilash ana shu farqlarni tizimli ravishda tahlil qilish imkonini beradi.

5. Anaforik razmetka. Anaforik razmetka matn tarkibidagi birliklar o‘rtasida yuzaga keladigan referensial munosabatlarni aniqlashga xizmat qiladi. Xususan, olmoshlar yoki boshqa ishoraviy birliklarning avvalgi matn qismida tilga olingan shaxs, predmet va hodisalarga munosabati belgilanadi.

Masalan, “Malika kutubxonaga bordi. U kerakli kitobni topdi” misolida *u* olmoshi avvalgi gapda tilga olingan Malikaga ishora qiladi. Anaforik razmetka ushbu birliklarning o‘zaro aloqasini belgilash orqali matnning mazmuniy yaxlitligini tahlil qilish imkonini beradi.

Bunday razmetka matnni avtomatik qayta ishlash, referentlarni aniqlash, matnni umumlashtirish, axborot ajratib olish va diskurs tahlilida muhim o‘rin tutadi. Shuningdek, u olmoshlarning qo‘llanishi, takroriy nomlash usullari va matn ichidagi bog‘lanishlarni o‘rganishda qo‘llanadi.

Anaforik munosabatlarni aniqlashda kontekst, grammatik moslashuv, semantik muvofiqlik va matnning umumiy mazmuni hisobga olinishi zarur. Ayrim hollarda bir olmosh bir nechta ehtimoliy referentga ishora qilishi mumkin, bu esa tahlilni murakkablashtiradi.

6. Prosodik razmetka. Prosodik razmetka, asosan, og‘zaki nutq korpuslarida qo‘llanib, nutqning ohang, urg‘u, pauza, ritm, davomiylik va boshqa prosodik xususiyatlarini belgilashga qaratiladi. Ayrim hollarda tovush balandligi, intensivlik va segmentlarning davomiyligi kabi akustik ko‘rsatkichlar ham qayd etiladi.

Prosodik xususiyatlar nutq mazmunining ifodalanishi va tinglovchi tomonidan qabul qilinishiga ta’sir ko‘rsatadi. Masalan, bir xil so‘zlar turlicha intonatsiya bilan aytilganda so‘roq, ta’kid, hayrat yoki boshqa kommunikativ ma’nolarni ifodalashi mumkin.

Prosodik razmetka og‘zaki nutqning fonetik tashkil etilishini o‘rganish, gaplarning kommunikativ turlarini farqlash, so‘zlovchining urg‘u orqali ajratgan birliklarini aniqlash va nutqni avtomatik tanish tizimlarini takomillashtirishda muhim ahamiyatga ega.

Ushbu razmetka turi yozma matnlarda to‘liq aks etmaydigan ko‘plab fonetik va kommunikativ xususiyatlarni qayd etish imkonini beradi. Shu sababli prosodik tahlil uchun matn bilan birga audioyozuv, vaqt belgilari va zarur hollarda akustik tahlil natijalari talab etiladi.

7. Diskursiv razmetka. Diskursiv razmetka matn yoki nutqning kommunikativ tashkil etilishi, mazmuniy qismlari o‘rtasidagi munosabatlar va birliklarning diskursdagi funksiyalarini belgilashga xizmat qiladi. Bunda sabab-natija, izohlash, qarama-qarshilik, davom ettirish, xulosa chiqarish va boshqa diskursiv munosabatlar qayd etilishi mumkin.

Masalan, “Yomg‘ir kuchaydi, shuning uchun tadbir bekor qilindi” jumlasida *shuning uchun* birikmasi sabab-natija munosabatini ifodalaydi. Diskursiv razmetka ushbu munosabatni maxsus teg yordamida belgilash imkonini beradi.

Diskursiv razmetka nafaqat alohida gaplarni, balki gaplar va matn qismlari o‘rtasidagi mazmuniy aloqalarni ham o‘rganishga yordam beradi. U matnning mantiqiy tuzilishini aniqlash, argumentatsiya usullarini tahlil qilish, kommunikativ strategiyalarni o‘rganish va matnni avtomatik umumlashtirish tizimlarini rivojlantirishda qo‘llanadi.

Diskursiv razmetka anaforik razmetka bilan uzviy aloqada bo‘lishi mumkin, biroq ularning vazifalari bir xil emas. Anaforik razmetka, asosan, referensial birliklar o‘rtasidagi munosabatlarni belgilasa, diskursiv razmetka matn qismlari o‘rtasidagi kengroq mazmuniy va kommunikativ aloqalarni qamrab oladi.

8. Razmetka turlarining o‘zaro aloqasi va statistik tahlildagi o‘rni. Lingvistik korpuslarda razmetka turlari alohida vazifalarni bajarsa-da, murakkab tadqiqotlarda ular o‘zaro uyg‘unlikda qo‘llanadi. Masalan, biror grammatik shaklning matndagi vazifasini to‘liq tavsiflash uchun uning morfologik xususiyatlarini, gap tarkibidagi sintaktik munosabatlarini va kontekstdagi semantik funksiyasini birgalikda o‘rganish talab qilinishi mumkin.

Razmetkalangan korpuslar yordamida quyidagi tahlillarni amalga oshirish mumkin:

- muayyan grammatik shakl yoki so‘z turkumining qo‘llanish chastotasini aniqlash;
- leksik birliklarning semantik guruhlar bo‘yicha taqsimlanishini o‘rganish;
- sintaktik konstruksiyalarning turli janr va uslublarda qo‘llanishini qiyoslash;
- anaforik munosabatlarning matnlar bo‘yicha tarqalishini tahlil qilish;
- og‘zaki nutqdagi prosodik xususiyatlarni qayd etish va qiyoslash;
- diskursiv munosabatlarning matnning mantiqiy tuzilishidagi o‘rnini aniqlash.

Bunda statistik ko‘rsatkichlar lingvistik hodisalarning miqdoriy tomonini ochib beradi, kontekstual va funksional tahlil esa ularning mazmunini izohlashga xizmat qiladi. Shu sababli korpus tadqiqotlarida miqdoriy va sifat tahlilini birgalikda qo‘llash maqsadga muvofiqdir.

Razmetkalash natijalarining ishonchliligi teglar tizimining aniqligi, belgilash qoidalarining izchilligi va matnlarni qayta ishlash jarayonining nazorat qilinishiga bog‘liq. Avtomatik razmetkalash katta hajmdagi materiallarni tez qayta ishlash imkonini bersa-da, ko‘p ma’noli birliklar, murakkab sintaktik tuzilmalar va kontekstga bog‘liq munosabatlarda xatolar yuz berishi mumkin. Shu bois ilmiy tadqiqotlarda razmetka natijalarini tekshirish va zarur hollarda tuzatish muhim hisoblanadi.

Natijalar va muhokama. Tahlillar lingvistik razmetka korpus materiallarini oddiy elektron matnlar majmuasidan tizimli lingvistik ma’lumotlar bazasiga aylantirishda muhim ahamiyatga ega ekanini ko‘rsatadi. Har bir razmetka turi tilning muayyan sathini yoki matnning alohida xususiyatini tavsiflashga yo‘naltirilgan.

Morfologik razmetka grammatik shakllarni, semantik razmetka ma’no xususiyatlarini, sintaktik razmetka gap qurilishi va grammatik munosabatlarni aniqlashga xizmat qiladi. Anaforik razmetka matn ichidagi referensial aloqalarni, prosodik razmetka og‘zaki nutqning ohang va akustik belgilarini, diskursiv razmetka esa matn qismlari o‘rtasidagi mazmuniy va kommunikativ munosabatlarni o‘rganish imkonini beradi.

Ushbu razmetka turlarining o‘zaro uyg‘unligi ko‘p qatlamli korpuslarni shakllantirish uchun metodologik asos yaratadi. Bunday korpuslarda bir xil matn materiali bir nechta lingvistik mezon bo‘yicha tahlil qilinishi mumkin. Natijada tadqiqotchi til hodisalarining alohida xususiyatlarini emas, balki ularning o‘zaro munosabatlarini ham o‘rganish imkoniyatiga ega bo‘ladi.

O‘zbek tilshunosligida morfologik, sintaktik va semantik razmetka bilan bir qatorda anaforik hamda diskursiv belgilash vositalarini rivojlantirish muhim ilmiy-amaliy ahamiyatga ega. Ayniqsa, o‘zbek tilining morfologik xususiyatlarini hisobga oluvchi avtomatik teglash tizimlarini takomillashtirish, og‘zaki nutq korpuslarini prosodik ma’lumotlar bilan boyitish va turli janrlardagi matnlarni diskursiv tahlil qilish imkoniyatlarini kengaytirish istiqbolli yo‘nalishlar hisoblanadi.

Shu bilan birga, razmetka tizimlarini ishlab chiqishda yagona va izchil belgilash qoidalarini shakllantirish, teglardan foydalanish mezonlarini aniq belgilash hamda avtomatik tahlil natijalarini tekshirishga alohida e’tibor qaratish lozim. Bu omillar korpusdan olingan statistik ma’lumotlarning ishonchliligini oshirishga xizmat qiladi.

Xulosa. Lingvistik korpuslarda razmetkalash til birliklari va matn komponentlarining xususiyatlarini tizimli qayd etish, ularni avtomatik izlash va statistik tahlil qilish imkonini beruvchi muhim metodologik vositadir. Lingvistik hamda ekstralingvistik teglar korpus materiallarini turli mezonlar asosida tavsiflash va ilmiy tadqiqotlar uchun moslashtirishga xizmat qiladi.

Maqola doirasidagi tahlil asosida quyidagi xulosalarga kelindi:

1. Lingvistik razmetka matn tarkibidagi birliklarning grammatik, semantik, sintaktik, referensial, prosodik va diskursiv xususiyatlarini tizimli belgilash imkonini beradi.
2. Ekstralingvistik razmetka matnning muallifi, yaratilgan vaqti, janri, manbasi va kommunikativ sharoiti kabi ma’lumotlarni qayd etish orqali lingvistik natijalarni qo‘shimcha mezonlar asosida qiyoslashga yordam beradi.
3. Morfologik, semantik, sintaktik, anaforik, prosodik va diskursiv razmetka turli tadqiqot vazifalariga xizmat qiladi hamda til hodisalarini ko‘p sathli tahlil qilish imkoniyatini kengaytiradi.
4. Razmetka turlarini birgalikda qo‘llash grammatik, semantik va diskursiv hodisalar o‘rtasidagi munosabatlarni aniqlash uchun muhim sharoit yaratadi.
5. Korpus asosida olingan statistik natijalarning ishonchliligi matnlarning reprezentativligi, teglar tizimining aniqligi va belgilash qoidalarining izchilligiga bog‘liq.
6. O‘zbek tili korpuslarini turli razmetka turlari bilan boyitish milliy tilning grammatik va semantik xususiyatlarini chuqur o‘rganish, avtomatik tahlil tizimlarini rivojlantirish va qiyosiy tadqiqotlarni kengaytirish uchun muhim istiqbollarni ochadi.

Umuman olganda, lingvistik korpuslarda razmetkalash til hodisalarini empirik asosda o‘rganish, tadqiqot natijalarini statistik dalillar bilan boyitish va tilshunoslikning zamonaviy axborot texnologiyalari bilan integratsiyasini ta’minlashga xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR

1. McEnery, T., Hardie, A. *Corpus Linguistics: Method, Theory and Practice*. – Cambridge: Cambridge University Press, 2012.

2. Biber, D., Conrad, S., Reppen, R. *Corpus Linguistics: Investigating Language Structure and Use*. – Cambridge: Cambridge University Press, 1998.

3. Jurafsky, D., Martin, J. H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd ed. – Upper Saddle River, NJ: Prentice Hall, 2009.

4. Sinclair, J. *Corpus, Concordance, Collocation*. – Oxford: Oxford University Press, 1991.

5. McEnery, T., Wilson, A. *Corpus Linguistics*. 2nd ed. – Edinburgh: Edinburgh University Press, 2001.