



MATNSHUNOSLIKDA KORPUS LINGVISTIKASI VA RAQAMLI TEXNOLOGIYALAR ASOSIDA MATN TAHLILINING ZAMONAVIY ILMIY YONDASHUVLARI

ISMATULLAYEVA AZIZA MIRZOHID-QIZI

Osiyo xalqaro universiteti 1-kurs magistri

Ona tili va adabiyot

ANNOTATSIYA

Ushbu maqolada matnshunoslikda korpus lingvistikasi va raqamli texnologiyalar asosida matn tahlilining zamonaviy ilmiy yondashuvlari nazariy hamda metodologik jihatdan tahlil qilingan. Tadqiqotda korpus lingvistikasi metodlari, tabiiy tilni qayta ishlash texnologiyalari, avtomatlashtirilgan lingvistik tahlil, statistik modellashtirish va kompyuter lingvistikasining matnning leksik-semantik, morfologik, sintaktik hamda pragmatik xususiyatlarini aniqlashdagi ilmiy imkoniyatlari yoritilgan. Shuningdek, elektron matn korpuslari asosida til birliklarini empirik tahlil qilish, lingvistik qonuniyatlarni aniqlash va matnni kompleks tadqiq etishda raqamli texnologiyalarning metodologik ustunliklari asoslab berilgan.

Kalit soʻzlar: *matnshunoslik, korpus lingvistikasi, raqamli texnologiyalar, elektron matn korpusi, kompyuter lingvistikasi, tabiiy tilni qayta ishlash, avtomatik lingvistik tahlil.*

XX asrning ikkinchi yarmidan boshlab dunyo tilshunosligida matnni empirik ma'lumotlar asosida o'rganishga bo'lgan ehtiyoj keskin ortdi. An'anaviy tilshunoslikda matn tahlili ko'pincha olimning individual kuzatuvi, intuitiv fikrlashi yoki cheklangan hajmdagi misollar asosida amalga oshirilgan bo'lsa, kompyuter texnologiyalarining rivojlanishi natijasida millionlab soʻzlardan iborat elektron



matnlar bazasini yaratish imkoniyati paydo bo'ldi. Aynan shu jarayon korpus lingvistikasining mustaqil ilmiy yo'nalish sifatida shakllanishiga asos soldi.

Korpus lingvistikasining zamonaviy tarixida muhim burilish nuqtasi 1961-yil hisoblanadi. Aynan shu yilda AQShning Brown universiteti olimlari W. Nelson Francis va Henry Kučera rahbarligida dunyodagi birinchi kompyuterda o'qiladigan umumiy til korpusi – Brown Corpus yaratildi. Ushbu korpus 500 ta matndan iborat bo'lib, umumiy hajmi taxminan 1 million so'zni tashkil etdi. Korpus tarkibiga gazeta matnlari, ilmiy maqolalar, badiiy adabiyotlar, rasmiy hujjatlar va publitsistik matnlar kiritildi. Ushbu loyiha til birliklarini statistik usullar asosida tahlil qilish imkonini yaratdi va keyingi korpuslar uchun metodologik namuna vazifasini bajardi.

1964-yilda Brown Corpus uchun rasmiy metodik qo'llanma nashr qilinishi korpus metodining ilmiy standartlarini belgilab berdi. Unda matnlarni tanlash mezonlari, korpus hajmi, janrlar bo'yicha taqsimlash tamoyillari hamda statistik ishlov berish usullari batafsil bayon qilindi. Natijada til hodisalarini alohida misollar orqali emas, balki katta hajmdagi real matnlar asosida tadqiq etish ilmiy jihatdan ishonchli metod sifatida tan olina boshladi.

1970-1980-yillarda kompyuter texnologiyalarining tez rivojlanishi korpus lingvistikasining yangi bosqichini boshlab berdi. Ushbu davrda Buyuk Britaniyada LOB Corpus (Lancaster-Oslo/Bergen Corpus), keyinchalik COBUILD Corpus, Bank of English kabi yirik elektron korpuslar yaratildi. Ushbu korpuslar matnlarning nafaqat leksik tarkibini, balki grammatik, sintaktik va uslubiy xususiyatlarini ham avtomatik tahlil qilish imkoniyatini yaratdi. Aynan shu davrda konkordans, kollokatsiya va chastota tahlili lingvistik tadqiqotlarning asosiy metodlariga aylandi.

1980-yillarning oxirlarida korpus lingvistikasi alohida ilmiy yo'nalish sifatida shakllandi. Ingliz tilshunosi John Sinclair korpus metodining nazariy asoslarini ishlab chiqdi hamda 1991-yilda nashr etilgan *Corpus, Concordance, Collocation* asarida real matnlar asosida tilni o'rganish intuitiv yondashuvdan ancha ishonchli ekanligini ilmiy dalillar bilan asoslab berdi. Sinclair til birliklari alohida emas, balki



kontekst ichida o'rganilgandagina ularning haqiqiy funksiyasi namoyon bo'lishini ta'kidlagan. Ushbu nazariya zamonaviy matnshunoslikning metodologik asoslaridan biriga aylandi.

1990-yillarda korpus lingvistikasining rivojlanishi bilan bir qatorda matnshunoslik ham yangi bosqichga ko'tarildi. Oldin matn faqat grammatik yoki stilistik birlik sifatida talqin qilingan bo'lsa, korpus texnologiyalari yordamida matnning semantik, pragmatik va diskurs xususiyatlarini millionlab misollar asosida tekshirish imkoniyati yuzaga keldi. Bu esa matn tahlilining obyektivligini sezilarli darajada oshirdi.

1998-yilda Douglas Biber, Susan Conrad va Randi Reppen tomonidan chop etilgan *Corpus Linguistics: Investigating Language Structure and Use* asari korpus lingvistikasining ilmiy metodologiyasini yanada rivojlantirdi. Mualliflar matnlarni statistik tahlil qilish orqali grammatik konstruktsiyalar, uslubiy farqlar va diskurs xususiyatlarini aniqlash mumkinligini ko'rsatdilar. Mazkur yondashuv matnshunoslikda sifat va miqdoriy metodlarning integratsiyasiga asos soldi.

2000-2015-yillar oralig'ida raqamli texnologiyalar rivojlanishi bilan korpus lingvistikasi tabiiy tilni qayta ishlash (Natural Language Processing – NLP) yo'nalishi bilan chambarchas bog'lana boshladi. Bu davrda morfologik teglash (POS-tagging), sintaktik daraxtlar (parsing), semantik annotatsiya va avtomatik lemmatizatsiya kabi texnologiyalar keng joriy qilindi. Natijada matn tahlili faqat so'z chastotasini aniqlash bilangina cheklanmay, matnning ichki semantik aloqalarini ham avtomatik ravishda tahlil qilish imkoniyati paydo bo'ldi.

XXI asrning ikkinchi o'n yilligidan boshlab sun'iy intellekt texnologiyalarining rivojlanishi korpus lingvistikasiga yangi bosqich olib kirdi. Ayniqsa, 2020-yil 28-mayda GPT-3 modeli taqdim etilgach, milliardlab so'zlardan tashkil topgan korpuslar asosida chuqur neyron tarmoqlarni o'qitish mumkinligi amalda isbotlandi. Bu hodisa matn tahlilida statistik yondashuvdan semantik va kontekstual tahlilga o'tishning yangi bosqichini boshlab berdi. Endilikda sun'iy



intellekt nafaqat soʻzlarning uchrash chastotasini, balki ularning yashirin semantik bogʻlanishlarini ham aniqlash imkoniyatiga ega boʻldi.

Soʻnggi yillarda korpus lingvistikasining rivojlanishi oʻzbek tilshunosligida ham jadallashdi. 2022-yil 27-oktabrda oʻzbek tilining morfologik va sintaktik teglangan korpusini yaratishga bagʻishlangan ilmiy tadqiqot eʼlon qilinib, unda POS teglash tizimi, sintaktik annotatsiya va maxsus veb-platforma ishlab chiqilgani bayon etildi. Mazkur loyiha oʻzbek tili uchun zamonaviy NLP texnologiyalarini yaratish yoʻlidagi muhim qadamlardan biri boʻldi.

2025-yilda Oʻzbekiston Milliy universiteti olimlari tomonidan chop etilgan “Korpus lingvistikasi rivojlanishi va korpus metodining nazariy asoslari” maqolasida korpus lingvistikasining tarixiy rivojlanish bosqichlari, elektron korpuslarning turlari, sinxron va diaxron korpuslar, parallel korpuslar hamda annotatsiya tizimlarining zamonaviy lingvistik tadqiqotlardagi oʻrni batafsil yoritildi. Tadqiqotda korpus metodining matnshunoslikdagi ilmiy ishonchlilik darajasini oshirishi va natijalarning takrorlanuvchanligini taʼminlashi alohida taʼkidlangan.

Shuningdek, 2025-yilda chop etilgan “Matnlarni lingvostatistik tahlil qilishda korpus usullaridan foydalanish” nomli ilmiy maqolada korpus lingvistikasining xronologik matn tahlilidagi ahamiyati chuqur tahlil qilinadi. Muallif korpus texnologiyalari yordamida til birliklarining tarixiy evolyutsiyasini, leksik oʻzgarishlarni, kollokatsiyalarni va uslubiy dinamikani aniqlash mumkinligini ilmiy misollar asosida asoslaydi. Tadqiqot natijalari matnshunoslikda empirik va statistik metodlarning ustuvor ahamiyat kasb etayotganini koʻrsatadi.

Bugungi kunda korpus lingvistikasi matnshunoslikning eng muhim ilmiy metodlaridan biri hisoblanadi. Zamonaviy matn tahlili faqat filologik kuzatishlarga emas, balki elektron korpuslar, sunʼiy intellekt algoritmlari, tabiiy tilni qayta ishlash texnologiyalari va katta hajmdagi maʼlumotlar tahliliga tayangan holda olib



borilmoqda. Shu sababli korpus lingvistikasi matnshunoslikni raqamli transformatsiyaga olib kirgan eng muhim ilmiy yo‘nalish sifatida e’tirof etiladi.

XULOSA

Raqamli texnologiyalar matnshunoslikning metodologik bazasini tubdan yangiladi. Elektron korpuslar, avtomatik annotatsiya tizimlari, morfologik va sintaktik teglash, konkordans dasturlari hamda sun'iy intellekt asosidagi til modellari ilmiy tadqiqotlarning samaradorligini oshirib, katta hajmdagi matnlarni qisqa vaqt ichida chuqur tahlil qilish imkonini bermoqda. Bu esa nafaqat nazariy tilshunoslik, balki amaliy tilshunoslik, tarjimashunoslik, kompyuter lingvistikasi, axborot qidiruv tizimlari va ta'lim texnologiyalarining rivojlanishiga ham xizmat qilmoqda.

O‘zbek tilshunosligida ham korpus lingvistikasi bo‘yicha olib borilayotgan ilmiy izlanishlar so‘nggi yillarda jadallashdi. O‘zbek tili milliy korpusini yaratish, morfologik va sintaktik teglash tizimlarini ishlab chiqish hamda NLP texnologiyalarini joriy etish bo‘yicha amalga oshirilayotgan tadqiqotlar milliy til resurslarini raqamlashtirish va xalqaro ilmiy maydonga integratsiyalashuv uchun muhim asos yaratmoqda. Shu bilan birga, mavjud korpuslarni kengaytirish, turli uslub va janrdagi matnlar bilan boyitish, semantik annotatsiyani takomillashtirish hamda sun'iy intellektga asoslangan avtomatik tahlil vositalarini rivojlantirish dolzarb ilmiy vazifalar bo‘lib qolmoqda.

FOYDALANILGAN ADABIYOTLAR RO‘YXATI

1. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
2. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
3. McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.



4. O‘zbekiston Milliy universiteti. (2025). Korpus lingvistikasi rivojlanishi va korpus metodining nazariy asoslari. Acta NUUz ilmiy jurnali.
<https://journals.nuu.uz/index.php/actanuuz/article/view/10347>