

O‘ZBEK TILI MILLIY KORPUSI YARATILISHINING DOLZARBLIGI**Karimova Gulnoza Abdujalil qizi***Samarqand davlat universiteti O‘zbek tilshunosligi kafedrasida, ilmiy-tadqiqotchisi**E-mail: gulnozkarimova106@gmail.com*

Annotatsiya. *Mazkur maqolada so‘nggi yillarda o‘zbek tilshunosligida yil sayin rivojlanib borayotgan yangi yo‘nalish – korpus lingvistikasining ahamiyati, tavsifi, tadqiq darajasi, o‘rganilishi, olib borilayotgan tadqiqotlar haqida umumiy fikr-mulohazalar beriladi.*

Tayanch so‘zlar. *Kompyuter, korpus lingvistikasi, korpus, lug‘at, milliy korpus, lingvistik ta‘minot, tabiiy tilni qayta ishlash.*

Annotation. *This article presents general reflections on the significance, description, level of research, and study of a new direction that has been rapidly developing in recent years in Uzbek linguistics — corpus linguistics. It also discusses the ongoing research in this field.*

Keywords. *Computer, corpus linguistics, corpus, dictionary, national corpus, linguistic support, natural language processing.*

Аннотация. *В данной статье представлены общие размышления о значении, характеристике, степени изученности и исследовании нового направления, которое в последние годы стремительно развивается в узбекском языкознании — корпусной лингвистики. Также рассматриваются проводимые исследования в этой области.*

Ключевые слова. *Компьютер, корпусная лингвистика, корпус, словарь, национальный корпус, лингвистическое обеспечение, обработка естественного языка.*

XXI asrning boshlarida axborot texnologiyalari jadal sur‘atlar bilan rivojlanishi natijasida tilshunoslikda yangi yo‘nalish — korpus lingvistika shakllandi. Ushbu yo‘nalishning asosiy maqsadi tabiiy tildagi matnlarni katta hajmda to‘plash, elektron shaklda saqlash va ular ustida avtomatik tahlil olib borishdan iboratdir. Korpus lingvistika tilshunoslikdagi an‘anaviy kuzatuv va tavsifiy usullarni yangi bosqichga olib chiqdi, chunki bu yo‘nalish kompyuter texnologiyalariga asoslangan holda til hodisalarini obyektiv tahlil qilish imkonini beradi.

Korpus lingvistika (Corpus Linguistics) — bu tilni real matnlar asosida o‘rganishga qaratilgan ilmiy yo‘nalish bo‘lib, u til birliklarini statistik, kompyuter va tahliliy usullar yordamida o‘rganishni nazarda tutadi. “Korpus” atamasi lotincha corpus — “tanasi”, “majmua” degan ma‘noni anglatadi. Til korpusi esa ma‘lum bir tilga oid yozma va og‘zaki matnlarning elektron bazasidir.

Korpus lingvistikasi XX asrning ikkinchi yarmida shakllanib, dastlab ingliz tilshunosligida qo'llanila boshladi. 1960-yillarda AQSh va Buyuk Britaniyada yaratilgan Brown Corpus, London-Lund Corpus kabi loyihalar bu sohaning ilmiy poydevorini yaratdi. Keyinchalik British National Corpus (BNC), COCA (Corpus of Contemporary American English), German Reference Corpus (DeReKo) kabi yirik milliy korpuslar paydo bo'ldi. Ular turli tillarda leksik, grammatik va sintaktik tizimlarni ilmiy asosda o'rganishga keng yo'l ochdi.

Bugungi kunda korpus lingvistika nafaqat lingvistik tadqiqotlarda, balki ta'lim, tarjima, sun'iy intellekt, nutqni aniqlash va avtomatik tarjima tizimlarida ham keng qo'llanilmoqda. Shu bois, o'zbek tilshunosligida ham so'nggi yillarda korpus lingvistika va kompyuter lingvistikasiga e'tibor sezilarli darajada oshdi. Ayniqsa, O'zbek tili milliy korpusining yaratilishi bu sohada yangi bosqichni boshlab berdi. Ushbu korpus o'zbek tilining lug'aviy, grammatik, semantik va uslubiy xususiyatlarini raqamli shaklda tadqiq etish imkonini yaratmoqda. O'zbek kompyuter lingvistikasi, tabiiy tilni qayta ishlash (NLP), mashinali ta'lim (machine learning), ma'lumot qidiruvi (Data Mining) kabi yangi sohalar kesimida rivojlanmoqda hamda ushbu sohalarida erishilayotgan ilmiy-amaliy natijalar bir-birini to'ldirib bormoqda. O'zbekiston Respublikasi Prezidentining 2019-yil 21-oktabrdagi "O'zbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan oshirish chora-tadbirlari to'g'risida"gi PF-5850-son Farmoniga muvofiq davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga integratsiyallashuvini ta'minlashda qator ustuvor vazifalar ko'rsatib o'tildi. O'zbek tilining mavqeini mustahkamlash va uni til siyosati darajasida nufuzini oshirishda tilning raqamli resursini yaratishga doir qator dasturlar ishlab chiqilmoqda.

O'zbek tilshunosligida korpus lingvistika yo'nalishidagi izlanishlar 2000-yillarning boshlarida boshlangan. Bu davrda tilshunos olimlar M. Mirtojiev, N. Rustamov, D. To'laganova, Sh. Sodiqov, N. Abdurahmonova va boshqalar o'zbek tilini elektron shaklda qayta ishlash masalalari ustida tadqiqotlar olib bordilar. Ularning ishlarida o'zbek tili leksikasi, morfologiyasi va sintaksisini kompyuter texnologiyalari yordamida tahlil qilish imkoniyatlari o'rganildi.

So'nggi yillarda bu yo'nalish yangi bosqichga ko'tarilib, O'zbek tili milliy korpusi (OTMK) yaratilmoqda. Mazkur loyiha O'zbekiston Respublikasi Fanlar akademiyasi, O'zbekiston Milliy universiteti va boshqa ilmiy markazlar hamkorligida olib borilmoqda. Korpus tarkibiga adabiy, ilmiy, publitsistik, badiiy va og'zaki matnlar kiritilib, ular maxsus kodlash va teglash tizimi orqali tizimlashtiriladi.

Korpus lingvistikasida erishilgan natijalar kompyuter lingvistikasi va tabiiy tilni qayta ishlash sohaları uchun ham til texnologiyasi borasida tub burilishlar yasadi. Bunda tabiiy tilning mashina tilini yaratish, til bo'yicha statistik ma'lumotga ega bo'lish, sun'iy intellektning lisoniy va nutq modellarini shakllantirish, tilning leksikografik mashina fondini yaratish kabi qator amaliy ishlar uchun o'rganish

obyekti bo‘lib xizmat qiladi. Korpus lingvistikasi turli fanlar uchun o‘rganish obyekti va vositasi sifatida foydalanilmoqda. Mazkur sohada Respublikamizning bir qator oliy ta‘lim muassasalari, shuningdek, ilmiy tadqiqot institutlarida mazkur yo‘nalishlar bo‘yicha ilmiy izlanishlar olib borilmoqda. O‘zbek korpus lingvistikasida B.Mengliyev, Sh.Shahobiddinova, Z.Xolmanova, S.Karimov, L.Raupova, Sh.Hamroyeva, N.Abduraxmonova, G.Toirova, S.Umirova, M.Tursunov, B.Xolmuhammedov, J.Djumabayeva, G.Ergasheva, A.Eshmo‘minov, Sh.Gulyamovalarning tadqiqotlarini qayd etish o‘rinli.

Kompyuter lingvistika — bu tilni avtomatik tarzda tahlil qilish, sintaktik tuzilmani aniqlash, matnni tarjima qilish va nutqni tanib olishga qaratilgan ilmiy-amaliy sohadir. U korpus lingvistikaning nazariy asoslariga tayanadi va ularni dasturlash texnologiyalari bilan uyg‘unlashtiradi.

Hozirda o‘zbek tilida avtomatik tarjima tizimlari, morfologik tahlil dasturlari, nutqni aniqlash algoritmlari, chatbotlar, sun‘iy intellekt asosidagi til modellarini yaratish bo‘yicha amaliy ishlar olib borilmoqda. Bu jarayonlarda o‘zbek tili korpuslari asosiy ma‘lumot manbai sifatida xizmat qiladi.

Keyingi yillarda Toshkent davlat o‘zbek tili va adabiyoti universitetida faoliyat yuritayotgan tilshunos olimlar hamkorlikda V.Zaxarov, B.Mengliyev, Sh.Hamroyevalar sohada izlanish olib borayotgan tadqiqotchilar va talabalar uchun o‘quv qo‘llanma ishlab chiqdilar. “Korpus lingvistikasi” nomli ushbu o‘quv qo‘llanma Toshkentda 2023-yilda nashr etilib ilm ahliga e‘lon qilindi. Kitob korpus lingvistikasi nazariyasi, tashkil topishi, taraqqiyot yo‘li, va bugungi holati bilan tanishtiradi; korpus texnologiyalarini o‘rganish, korpuslar bilan ishlash ko‘nikmasini paydo qilish imkonini yaratadi. Mazkur adabiyot 185 sahifadan iborat bo‘lib, uch bobdan iborat: korpus lingvistikasiga kirish (bu bobda fanning maqsad va vazifalari, predmet va obyekti haqida, korpusning amaliy ahamiyati, undagi asosiy tushunchalar, til korpuslarining turlari va tipologiyasi haqida so‘z boradi), korpusdan foydalanish (mazkur bobda esa korpus lingvistikasi yo‘nalishining bugungi holati: jahon va turkiy tilshunoslikda, parallel korpuslar, mualliflik korpuslari haqida ma‘lumotlar beriladi), korpus tuzish texnologiyasi (so‘nggi bobda korpus tuzish texnologiyalari ya‘ni bosqichlar, razmetka texnologiyalari, tabiiy tilni qayta ishlash, lingvistik va ekstralingvistik razmetka, korpusning qidiruv tizimi ekanligi haqida axborot almashiniladi). Shu o‘rinda eslatib o‘tish joizki, korpus – elektron yig‘ilgan matnlar bazasidir. Korpus tarkibiga kiradigan matnlar quyidagi turlarda bo‘lishi mumkin:

- 1) alohida bir muallif yoki bir necha yozuvchi asaridan olingan matnlar;
- 2) ma‘lum davr (bir necha o‘n yillik yoki yuz yillik)ni qamrab olgan matnlar;
- 3) belgilangan mavzudagi zamonaviy matnlar;
- 4) til va jamiyatning bugungi holatini aks ettiruvchi matnlar.

O‘zbek tili milliy korpusi (OTMK) — bu o‘zbek tilidagi matnlarning raqamli bazasi bo‘lib, u til birliklarini tahlil qilish, yangi lug‘atlar tuzish, so‘zlarning

qo'llanish chastotasini aniqlash, kontekstual tahlil o'tkazish va til o'qitish jarayonini takomillashtirish imkonini beradi.

OTMK asosida tildagi semantik o'zgarishlar, yangi so'zlarning paydo bo'lishi, dialektal farqlar hamda uslubiy xususiyatlar ilmiy asosda o'rganiladi. Shu bilan birga, bu korpus tilni raqamlashtirish, milliy axborot tizimlarini yaratish, va ta'lim texnologiyalarini modernizatsiya qilish uchun ham muhim ahamiyat kasb etadi.

Korpus ma'lumotlari asosida yaratilgan dasturiy vositalar o'zbek tili grammatikasini o'rganishda, avtomatik tekshiruv tizimlarini yaratishda, shuningdek, o'zbek tilining xalqaro miqyosdagi o'rganilishini yengillashtirishda katta imkoniyatlar ochadi.

Korpus lingvistikasi va kompyuter lingvistikasi o'zbek tilshunosligining yangi bosqichini boshlab berdi. Ular tilni raqamli muhitda o'rganish, qayta ishlash va amaliy sohalarga tatbiq etish imkonini beradi. O'zbek tili milliy korpusi bu jarayonning markaziy elementi sifatida milliy tilni global ilmiy maydonga olib chiqishga xizmat qilmoqda. Kelgusida bu yo'nalishda avtomatik matn tahlili, sun'iy intellekt asosida lug'at tuzish, matnni tovushga aylantirish, nutqni tanish kabi texnologiyalarni yanada rivojlantirish o'zbek tilining xalqaro nufuzini oshirishga xizmat qiladi. Shu sababli, o'zbek tilining korpus lingvistikasi bo'yicha ilmiy izlanishlarni qo'llab-quvvatlash va uni amaliy sohalarga keng tatbiq etish muhim strategik ahamiyatga ega.

Foydalanilgan adabiyotlar

1. Adilova S. O'zbek tilini o'qitishda kompyuter texnologiyalaridan foydalanish. – Toshkent, 2006.
2. Muhamedova S. Kompyuter lingvistikasi (metodik qo'llanma). – Toshkent: ToshDPU nashri, 2007.
3. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017.
4. Раҳимов А. Компьютер лингвистикаси асослари. – Т.: Akademnashr, 2011.
5. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. –Иркутск: ИГЛУ, 2011.