

TIL KORPUSI VA UNING PREDMETI

Unarova Dilafruz Abdimajit qizi
SamDCHTI akademik litseyi
ona tili va adabiyot fani o'qituvchisi
dilafruzunarova1@g.mail.com

Annotatsiya. Ushbu maqolada til korpusi va uning predmeti xususida so'z boradi. Korpus lingvistikasining tilshunoslik fani bilan uzviy bog'liqligi va rivoji, tutgano'rni, amaliy ahamiyati tahlil va talqin etilgan.

Kalit so'zlar: Korpus haqida tushuncha, til birliklari, matn va korpus lingvistikasi, elektron kutubxona.

Korpus lingvistikasi va uning predmeti. Korpus lingvistikasi jahon kompyuter lingvistikasining juda tez rivojlanib ketayotgan sohasi bo'lib, bu borada qancha yutuqlarga erishilgan degan savolni qo'yadigan bo'lsak, albatta bu borada anch ishlarni ko'rishimiz mumkin. Dunyo oliy ta'lim muassasalarida korpus lingvistikasi fan sifatida ham o'qitilmoqda. Bu sohaning predmeti korpus yaratish nazariyasi va amaliyoti bo'lsa, fan sifatida korpusning o'ziga xosligi, dasturlash asoslari kabi jihatlari o'qitiladi. Korpus lingvistikasi kompyuter lingvistikasining tarkibiy qismi sifatida, til korpusini yaratish, kompyuter texnologiyasi yordamida ulardan foydalanishning umumiy nazariyasi va amaliyoti bilan shug'ullanadi deb ayta olishimiz mumkin. Korpus lingvistikasining asosiy predmeti— bu til korpusi hisoblanadi. U ingliz tilida linguistic corpus yoki text corpus kabi atamalar bilan qo'llanilib kelinmoqda, shuningdek, ilmiy adabiyotlarda turlicha ta'riflangan. “Korpus – ma'lum maqsadda yig'ilgan matnlar majmuini tashkil etuvchi til birliklarining yig'indisi”, “korpus – alohida tilning yozib olingan og'zaki hamda yozma matnlari asosidagi lingvistik ma'lumotlar majmuyi”, “korpus – istalgan tabiiy (real) tildagi elektron shaklda saqlanadigan yozma yoki og'zaki, kompyuterlashtirilgan qidiruv tizimiga joylashtirilgan matnlar yig'indisi” kabi. Demak, til korpusi ma'lum

tilning belgilangan davrdagi, xilma-xil janr, rang-barang uslub, hududiy hamda ijtimoiy variantdagi matnlarining elektron shakldagi maxsus dasturiy ta'minot asosidagi yig'indisidir. Korpus haqida yuqorida keltirilgan ta'riflar ham u haqida muayyan tilning o'ziga xos xususiyati va variantlarini aks ettiruvchi bir necha belgi asosida tanlab olingan elektron shakldagi matn parchalari, lingvistik tadqiqot uchun asos vazifasini o'taydigan tizim deya xulosa qilishga asos bo'ladi. Bu esa har bir ilmiy izlanuvchi uchun asos bo'lib xizmat qiladi. Til korpusi bugungi kunda jahonda tez taraqqiy etayotgan soha – korpus lingvistikasi mutaxassislari tomonidan ilmiy tadqiqotlar olib borish, til o'rgatish maqsadida yaratilmoqda. Jahon tillarining juda ko'pchiligi mukammallik darajasi, matnni (ilmiy) qayta ishlash imkoniyati bilan farq qiluvchi o'z milliy korpuslariga ega. Korpus oddiy elektron kutubxona bo'lib, (mas., “Библиотека Максима Мошкова”, “Русская виртуальная библиотека”, “Фундаментальная электронная библиотека”, “ZIYONET kutubxonasi”, “Xurshid Davron kutubxonasi” kabilar)dan ancha osonligi va tezligi bilan farq qiladi. Bunday elektron kutubxonalar xalqning ijtimoiy-siyosiy, ma'naviy, iqtisodiy hayotini aks ettiruvchi badiiy hamda publitsistik asarlarni nisbatan to'laroq qamrab olishni maqsad qilgan bo'ladi⁵. Faqat bunda elektron kutubxona matnlari tilshunoslik nuqtayi nazaridan leksik, morfologik, grammatik, semantik belgilar asosida ishlov berilmaganligi sababli lingvistik tadqiqotlarda foydalanish uchun noqulaylik tug'diradi. Til korpusi esa elektron kutubxonadan farqli ravishda muayyan tilni tadqiq qilish, o'rganish va o'rgatish uchun zaruriy, foydali va qiziqarli matnlar to'planishini nazarda tutadi. Bunday matnlar esa nafaqat badiiy adabiyot namunasi, balki telefondagi oddiy suhbat yozuvi yoki ijara shartnomasi matni ham bo'lishi mumkin. Korpusni elektron kutubxonadan ajratib turuvchi birinchi omil undagi matnning xususiyati, qo'shimcha ma'lumot bilan boyitilganligi hisoblanadi hamda bu korpusning alohida qismi – korpus annotatsiyasi (rus., разметка)da o'z ifodasini topadi. Korpus – ma'lum maqsadda yig'ilgan matnlar majmuini tashkil etuvchi til birliklari yig'indisi, tabiiy tildagi elektron shaklda saqlanadigan yozma yoki og'zaki, kompyuterlashtirilgan qidiruv tizimiga dasturiy ta'minot asosida joylashtirilgan on-line yoki off-line tizimda

ishlaydigan matnlar jamlanmasi. Lingvistik tadqiqot va amaliy topshiriqlar yechimi uchun til korpuslari zamonaviy tilshunoslikning inkor etib bo'lmaz ish quroli. U oddiy elektron kutubxonadan farqlanadi. Elektron kutubxona xalqning ijtimoiy-siyosiy, ma'naviy, iqtisodiy hayotini aks ettiruvchi badiiy va publisistik asarlarni nisbatan to'liq qamrab olishni maqsad qilgan bo'ladi. Elektron kutubxona matnlari tilshunoslik nuqtayi nazaridan leksik, morfologik, grammatik, semantik belgilar asosida ishlov berilmaganligi sababli lingvistik tadqiqot uchun noqulaylik tug'diradi. Chunki elektron kutubxona ilmiy tadqiqot materiali bazasini tayyorlash maqsadida tuzilmaydi, balki milliy ma'naviy merosni to'plashni maqsad qilgan bo'ladi. Til korpusi esa elektron kutubxonadan farqli o'laroq, tilni o'rganish uchun zaruriy, foydali va qiziqarli matnlarni to'plashni nazarda tutadi. Misol tariqasida biron bir asarda qancha so'zlar keltirilganligini bilish, shuningdek, ularning turkumlari, grammatik ko'rsatkichlari xususida ma'lumotlarga ega bo'lish imkonini beradi.

Foydalanilgan adabiyotlar ro'yxati

1. Хамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
2. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
3. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.
4. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.