

KORPUS TILSHUNOSLIGI HAQIDA MULOHAZALAR

Otamurodova Mohira Aziz qizi,

otamurodova-m@samdchti.uz

Samarqand davlat chet tillari instituti magistranti

Annotatsiya. Maqola zamonaviy tilshunoslikda korpus yaratish mahorati va uning o'ziga xos jihatlari haqida ma'lumot berilgan. Korpus tilshunosligining paydo bo'lishi, uning til taraqqiyotidagi vazifalari tahlil qilinadi.

Kalit so'zlar: *lingvistika, elektron kutubxona, korpus tilshunosligi, matn, ijtimoiy qatlam leksikasi, internet resurslari.*

Bugungi kunda yurtimizda o'zbek tiliga qaratilayotgan e'tibor nihoyatda katta. Bu esa tilshunoslarimiz, shuning bilan birga informatiklarimizning oldiga yana bir muhim vazifani qo'yadi-ki, u ham bo'lsa o'zbek tilining korpus lingvistikasini yaratishdir. Bugungi kunda dunyo tillari bo'lmish rus, ingliz, ispan, arab tili kabi barcha tillarda mavjud bo'lgan ushbu baza bizning tilshunosligimizda ham bor bo'lishini kutib turgan asosiy masaladir.

Korpus lingvistikasi-tilshunoslik kompyuter texnologiyalaridan foydalangan holda lingvistik ma'lumotlar korpusidan foydalanish hamda ularni yaratishning umumiy prinsiplarini ishlab chiqarish bilan shug'ullanadi.

Leksikaning korpus tavsifi leksikografiyada eng ko'p qo'llaniladi. Deyarli barcha zamonaviy ingliz lug'atlar korpuslar asosida yaratilgan. Korpus tilda so'zlar to'plamini ishonchli tarzda aniqlashga yordam beradi, yangi so'z shakllarining paydo bo'lishini ko'rsatadi, bir so'zning turli ma'nolari va uning chastotasini aniqlash uchun ishlatiladi.

Bu yerda og'zaki fe'lining shakllari, old qo'shimchalar, ravishlar va boshqalar haqida ham gapirish mumkin. Fe'llarni shakllarini o'rganish bilan shug'ullanadigan tadqiqotlar maxsus o'ringa ega, chunki Braun va LOB korpuslarini o'rganish natijalariga ko'ra, barcha so'zlarning deyarli 20 foizini tashkil etuvchi bu fe'llar, barcha so'z foydalanishlarning 95 foizini tashkil etuvchi 23 ta eng ko'p ishlatiladigan fe'llar mavjud.

Korpus til haqida juda qiziqarli leksikografik ma'lumotlar berishi mumkin. Korpuslar asosida leksikaning statistik tadqiqotlari kompyuterlardan oldin davrda boshlangan. Masalan, ingliz tilidagi matnlarda so'zning 90-95 foizi 2-5 mingga eng

keng tarqalgan soʻz shakllaridan iborat. Bundan tashqari, matnning taxminan yarmi soʻzning 50-100 ta eng mos keladigan soʻz shakllaridan foydalanishidir (garchi ularning turli xil uslublari boʻlishi mumkin). Bu esa ingliz tilini oʻqitish uchun katta ahamiyatga ega, chunki oʻquvchilarni eng tez-tez ishlatiladigan soʻzlar bilan tanishtirishga imkon beradi.

Maʼlumki, korpus tilshunosligi baʼzi tilshunoslar tomonidan tadqiqot vositasi yoki metodikasi, boshqalari esa oʻz navbatida intizom yoki nazariya sifatida koʻrib chiqiladi. Korpus tilshunosligida qoʻllaniladigan usullar ilk bor 1960-yillarning boshlarida qabul qilingan boʻlsa-da, korpus tilshunosligi 1980-yillarga qadar paydo boʻlmagan. Til korpuslari - til boʻyicha tadqiqot va amaliy topshiriqlar yechimi uchun inkor etib boʻlmay ish quroli. U oddiy elektron kutubxonadan farqlanadi. Elektron kutubxonaning maqsadi - xalqning ijtimoiy-siyosiy, maʼnaviy, iqtisodiy hayotini aks ettiruvchi badiiy va publitsistik asarlarni nisbatan toʻliq qamrab olish. Elektron kutubxona matnlari til nuqtayi nazaridan ishlov berilmaganligi sababli tadqiqotlar uchun noqulaylik tugʻdiradi. Chunki, elektron kutubxona ilmiy tadqiqot materialini bazasini tayyorlash maqsadida tuzilmaydi, balki milliy maʼnaviy merosni jamlashni maqsad qilgan boʻladi. Til korpusi esa elektron kutubxonadan farqli oʻlaroq, tilni oʻrganish va tadqiq qilish uchun zarur, foydali va qiziqarli matnlarni toʻplashni nazarda tutadi.

Koʻpgina mamlakatlarda XX asrning 80-yillaridan boshlab bunday korpuslar tuzila boshlandi. Ular turli maqsad va vazifalarga xizmat qiladi. Buyuk Britaniyada Ingliz tili Banki (Bank of English) hamda Britaniya milliy korpusi (British National Corpus, BNC), Rossiyada Rus tilining milliy korpusi loyihalari ishlab chiqildi. Masalan, Rus tilining Milliy Korpusi hajmi hozirgi kunda 149 mln soʻzdan iborat. Keyingi yillarda Internet tizimining rivojlanishi virtual matnlar korpusi yuzaga kelishiga olib keldi. Yaʼni internetdagi qidiriv saytlari, elektronkutubxonalar, virtual ensiklopediyalar korpus vazifasini bajarmoqda. Korpusning janri va tematik rang-barangligi Internetdan foydalanuvchining qiziqishlariga bogʻliq. Oʻzbek tilida milliy korpus yaratish ishlari hanuz orqaga surilmoqda.

Til korpusi tilda yaratilgan turli tipdagi matnlar – adabiy-badiiy (mumtoz va zamonaviy adabiyot namunalari), ilmiy-ommabop, yozma va ogʻzaki, adabiy va noadabiy qatlamga oid, sheva va lahjalarni aks ettirgan, umumxalq tili boyligi jamlangan manbalarni qamrab oladi. Korpuslarning maqsadi bir qancha yangi faktlarni namoyish qilish emas, olingan natijalarni shunchaki aniqlab asoslab berishdir.

Tilshunoslik korpusi tilning strukturaviy va funktsional parametrlarini aniqroq va to'g'ri tavsiflash uchun asos yaratadi. Bugun biz ba'zi bir korpus tadqiqotlarining natijalari haqida gapiramiz va umuman olganda korpuslarning ishlatilishi tilshunoslikka qanday yordam bera olishi haqida gaplashamiz.

Zotan, korpus tushunchasi zamonaviy lingvistikaning asosiy tayanch tushunchasidir. Til korpusi muayyan bir tildagi badiiy, ilmiy, publitsistik va boshqa turdagi matnlarning qidiruv dasturiga bo'ysundirilgan elektron shakldagi majmuidir. Shuni ham qayd etish joizki, til korpusi tilda yaratilgan turli tipdagi matnlar – adabiy-badiiy (mumtoz va zamonaviy adabiyot namunalari), ilmiy-ommabop, yozma va og'zaki, adabiy va noadabiy qatlamga oid, sheva va lahjalar, ijtimoiy qatlamlar leksikasini aks ettirgan, umumxalq tili boyliklari jamlangan manbalarni qamrab oladi.

Til korpuslarining hajmi turlicha bo'lib, ularda 1000-2000 so'zdan tortib, millionlab birikma va iboralar o'z aksini topgan bo'lishi mumkin. Til korpusi nihoyatda qisqa muddatda kerakli til dalilini izlash va tahlil qilish imkoniyatini beradi. Lug'at va grammatikaning korpus tarkibiga kirishi, ulardagi maqolalarning ishonarli, ob'ektiv bo'lishi va har kuni to'ldirib borilishi til korpusining eng muhim faktorlaridandir. Korpus internet matnlarini avtomatik tanlash, matn qurilishini avtomatik tahlil qilish, avtomatik kompyuter xotirasida saqlash(ovozli korpuslar yordamida) va boshqa vazifalarni bajara oladi. Internetda korpusli portallarning tashkil etilishi ijtimoiy-ta'limiy ahamiyat kasb etadi. Til korpuslariga asoslangan internet resurslari tilshunos bo'lmagan turli soha vakillarining til va uning tarixi (so'zlar, ularning tarixiy yasalishi, etimologiyasi, fonetikasi, orfagrafiyasi) bilan bog'liq savollarga javob beradi[3].

Ko'rinadiki, korpuslar katta hajmdagi ma'lumotlarni qayta ishlashga imkon beradi. Ular yordamida ikki yoki undan ko'p tilli lug'atlar tuzish mumkin. Gaplarni qiyosiy uzunligi, matnlarni abzatslarga bo'linishi, imlo belgilarini tahlillab, tashqi lug'atdan foydalanib, boshqa faktorlar asosida ham maqsad qilingan vazifalarni bajarish mumkin. Ko'pincha bu holatlarda dasturlar odam va mashina variantida ishlatiladi.

Elektron korpuslar bizga moslik muammosini ob'ektiv ravishda ko'proq o'rganishga imkon beradi. Masalan, biz korpusdagi so'zlarning aslida birlashtirilishi va ularning individual chastotlariga asoslangan holda nazariy jihatdan qanday qilib birlashtirilishi kerakligi o'rtasidagi farqning statistik ahamiyatini hisoblashimiz mumkin. Masalan, Christmas so'zi odatda day, eve, tree, cards va present kabi so'zlar bilan birlashtiriladi.



Korpuslarni tahlil qilishdagi keyingi yangiliklar tilni tavsiflash sohasida tadqiqotlar uchun katta maydonni ochib beradi. Bu yerda korpus lingvistikasi leksika va grammatika orasidagi chegarani aniqlashga yordam beradi. Zamonaviy kompyuterlarning imkoniyatlari-xotirasi, takt chastotasi, tashqi qurilmalardan foydalanish imkoniyatlari, matnlarni skanerlash, kompakt-disklarga, fleshka kartalarga yozish, dasturlovchilarning katta masshtabdagi korpuslar bilan bajariladigan ishlarni ta'minlaydi. Bu esa, ko'zlangan maqsadni amalga oshirish uchun muhim hisoblanadi.

Foydalanilgan adabiyotlar:

1. Yuldashev B. Kompyuter lingvistikasi, Samarqand.2011.
2. Po'latov A.Q.Kompyuter lingvistikasi,Samarqand.2011.
3. V.P.Zaxarov, S.Y.Bogdanova. Корпусная лингвистика. Irkutsk,2011.
4. Turayev H.T.Matematik mantiq va diskret matematika.Samarqand,2003.