MILLIY TIL KORPUSI VA UNING AHAMIYATI

Gulnoza Karimova,
Sharof Rashidov nomidagi Samarqand davlat universiteti, Markaziy Osiyo xalqlari tillari va madaniyati instituti doktoranti

Annotatsiya. Ushbu tezisda jahon va o‘zbek tilshunosligida korpus masalasining o‘rganilishi, bu borada amaliy ishlar ham tadqiq qilinganligi haqida so‘z yuritilgan. Shuningdek, tezisda o‘zbek tili milliy korpusining yaratilishi bugunning eng global zarurati ekanligi ham yoritib berilgan.

Kalit so‘zlar: *milliy korpus, kompyuter lingvistikasi, til siyosati, raqamlashtirish, elektron matn, baza, texnika, integratsiyalashuv, ommalashtirish*.

Korpus biror bir xalqning ma’naviy xotirasini o‘zida aks ettiruvchi noyob texnologik ijoddir. *Milliy korpus* esa o‘zbek xalqining milliy tili va o‘zligini namoyon etadigan, o‘zbek tilini jahonga olib chiqadigan katta imkoniyat, ayni paytda yuksak zaruriyat.

XXI asr — texnika taraqqiyoti asri. Bugun zamonaviy axborot texnologiyalari ko‘plab sohalarda, xususan, tilning barcha foydalanish mumkin bo‘lgan obyektlari uchun ham cheksiz imkoniyatlar eshigini ochgani sir emas. O‘zbek tiliga davlat tili maqomining berilishi tilimizning milliy imkoniyatlari doirasini kengaytirdi. **O‘zbekiston Respublikasi Prezidenti Sh.Mirziyoyevning** 2020-yil 20-oktabrdagi “Mamlakatimizda o‘zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to‘g‘risida”gi farmoni bilan tasdiqlangan “2020–2030-yillarda o‘zbek tilini rivojlantirish va til siyosatini takomillashtirish konsepsiyasi”da davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga faol integratsiyalashuvini ta’minlash, ustuvor yo‘nalishidagi o‘zbek tiliga oid barcha

ilmiy, nazariy va amaliy ma’lumotlarni o‘zida jamlagan elektron ko‘rinishdagi o‘zbek tili milliy korpusini yaratish va o‘zbek tilini internet jahon axborot tarmog‘ida ommalashtirish, unda munosib o‘rin egallashini ta’minlash vazifalari belgilangani tilimiz uchun juda muhim jarayondir.[1]

Jahon tilshunosligi alohida soha sifatida rivojlandi, fan sifatida yuksak natijalarga erishdi. So‘nggi yillarda tilshunoslikda shunday muammolar tizimi yuzaga keldiki, ularni boshqa sohalar ishtirokisiz hal qilish mumkin bo‘lmay qoldi. Natijada tilshunoslik integratsiyasi asosida bir necha tutash sohalar yuzaga keldi. “Shu paytgacha, turli fan sohalari til tizimini, lisoniy qobiliyatni o‘z ehtiyojlari uchun, o‘z qiziqishi doirasida, turli tahlil uslublari vositasida tadqiq etib keldi. …Lison muammosi bilan qiziquvchi fanlar... yagona bir ob’ektni majmuaviy, har tomonlama tadqiq etish yo‘lini tanladi va natijada sohalararo fanlar, yo‘nalishlar shakllandi”.[2]

Kompyuter lingvistikasi (matematik lingvistika, hisoblash lingvistikasi) – tilshunoslik, matematika va informatika kesimida shakllangan bo‘lib, dasturiy ta’minot va kompyuter texnologiyalaridan foydalanishning tilshunoslik bilan bog‘liq, deyarli, barcha qirralarini qamrab oladi. Amaliy lingvistika sohasi sifatida qaralayotgan kompyuter lingvistikasi ijtimoiy ehtiyoj talabi bilan yuzaga keldi.

Korpus lingvistikasi esa jahon kompyuter lingvistikasining juda tez rivojlanib ketayotgan sohasi bo‘lib, uning taraqqiyotini belgilovchi asosiy ta’minot elektron matnlar korpusidir. Jahon tilshunosligida korpus lingvistikasi o‘tgan asrning 60- yillarida o‘rganish obyektiga aylangan edi. Matnlarni avtomatik tarjima qilish haqidagi g‘oyalarning amaliyotga tatbiq qilinishi, hisoblash lingvistikasi, matematik lingvistika asoslarining yuzaga kelishi ma’lumotlar bazasini shakllantirish ehtiyojini yuzaga keltirdi. Avtomatik tarjimani amalga oshirish uchun birinchi navbatda tarjima qilish ko‘zda

1 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 Farmoni (https://lex.uz/docs/-5058351)
[2] Сафаров Ш. Когнитив тилшунослик. Жиззах, 2006.

tutilayotgan tillardagi ma’lumotlar bazasi zarur edi. Korpus lingvistikasi masalalari ilk bor avtomatik tarjimani amalga oshirish g‘oyasiga muvofiq ravishda o‘rtaga tashlangan. Korpus haqidagi ilk fikrlar *R.G.Piatrovskiy* tomonidan aytilgan[3] bo‘lsa-da, bu sohadagi maqsadli tadqiqotlar 40-yillarda *Blumfild, Frays va Bondjerslar* tomonidan boshlangan[4]. *N.Frensis va G.Kuchera* tomonidan korpus tuzish prinsiplarining ishlab chiqilishi, *J.Sinkler* tomonidan dastlabki ingliz korpusining yaratilishi[5] korpus lingvistikasida ilk nazariy va amaliy qadam bo‘ldi. Rus tilshunosligida *V.P.Zaxarov, A.B.Kutuzov, Ye.V.Nedoshivina, V.V.Rikov, V.Plungyan, O.V.Kukushkina, A.A.Polikarpov, Ye.V.Surovtsevalar* korpus, uning turlari, o‘ziga xos xususiyati, korpusning ijtimoiy ahamiyati, korpus tuzish tamoyillari borasida tadqiqot olib borishgan.

Til korpuslari – til bo‘yicha tadqiqot va amaliy topshiriqlar yechimi uchun inkor etib bo‘lmas ish quroli. O‘zbek tilshunosligida ham kompyuter lingvistikasi, matnga leksikografik ishlov berish va lingvostatistik tahlil etish borasida tadqiqotlar amalga oshirilgan. *A.Po‘latov, S.Muhamedov, M.Ayimbetov, S.Muhamedova, S.Karimov, G.Jumanazarova, A.Babanarov, D.O‘rinboeva, N.Abdurahmonova, A.Norov* va boshqalarning tadqiqotlari ana shunday ishlar sirasiga kiradi[6]. Bu tadqiqotlar matnni kompyuter lingvistikasi yutuqlari yordamida leksikografik va lingvostatistik tadqiq etishning zamonaviy usullarini tavsiya etgan bo‘lsa-da, ularda o‘zbek lingvistik korpuslarini yaratish masalasi

3 Kutuzov A.B. Korpusnaya lingvistika. − (Elektron resurs): Litsenziya Creative commons Attribution Share.
4 Blumfild L. Yazыk. – M.: «Progress», 1968. – 608 s.; Fries Ch.C. The structure of English. An introduction to the construction of English sentences. – L.,1969.
5 Frensis N., Kuchera G. Vichislitelьnыy analiz sovremennogo amerikanskogo varianta angliyskogo yazika. – M., 1967.; Sinkler D. Predislovie k knige «Kak ispolzovat korpusa v prepodavanii inostrannogo yazik.
6 Muhammedov S.A. Statisticheskiy analiz leksiko-morfologicheskoy strukturы uzbekskix gazetnыx tekstov: Avtoref. diss… kand. fil. nauk.- Tashkent, 1980; Babanarov A. Razrabotka printsipov postroeniya slovarnogo obespecheniya turetsko-russkogo mashinnogo perevoda: Avtoref. diss… kand. fil. nauk. - L., 1981; Auыmbetov M.K. Opыt lingvostatisticheskogo analiza leksiki i morfologii karakalpakskogo publitsisticheskogo teksta: Avtoref. diss… kand. fil. nauk.- Tashkent, 1987; Karimov S., Qarshiev A., Isroilova G. Abdulla Qahhor asarlari tilining lug‘ati. Alfavitli lug‘at. Chastotali lug‘at. Ters lug‘at, - Toshkent, 2007; Rizaev S. O‘zbek tilining lingvostatistik tadqiqi: Fil.fan.dok.dis…avtoref. - Toshkent, 2008; Muhammedova S. Harakat fe’llari asosida kompьyuter dasturlari uchun lingvistik ta’min yaratish. Metodik qo‘llanma. - Toshkent, 2006; O‘rinboeva D.B. O‘zbek folklori matnlarining lingvostatistik tadqiqi. – Toshkent: «Fan», 2010; Jumanazarova G.U. Fozil Yo‘ldosh o‘g‘li dostonlari tilining lingvopoetikasi: Fil. fan. dok. dis…avtoref. - Toshkent, 2017.

kun tartibiga qo‘yilmagan. *Sh.Hamroeva* tomonidan olib borilgan tadqiqotni[7] mana shu yo‘nalishdagi amaliy qadam desa bo‘ladi. O‘zbek tilshunosligida bu borada qilingan ishlarni o‘zbek tili milliy korpusini yaratishdagi ilk qadam sifatida baholash mumkin. Til korpusi va unga oid ilmiy ishlar 2018-yildan boshlandi va bugunga qadar qator ilmiy yangiliklar qilinmoqda.

Axborot asrida rivojlanishga da’vogar har qanday til sun’iy intellekt tiliga aylanishi kerak. Jahon tillarining juda ko‘pchiligi mukammalik darajasi va matnni ilmiy qayta ishlash imkoniyati bilan farq qiluvchi o‘z milliy korpusiga ega. Ingliz tilida yaratilgan *Braun korpusi, Lankaster-Oslo/Bergen (LOB) korpusi, London-Lund korpusi,* Leksikografik tadqiqotlar uchun *Amerika meros korpusi (American Heritage Intermediate), Lankaster/IBM Spoken English Corpus, Corpus of Spoken American English, etc),* diaxronik korpus sanalgan Ingliz matnlarining *Xelsinki korpusi,* lingvodidaktik tadqiqotlar uchun *Ingliz tili o‘rganuvchilarining xalqaro korpusi*, ingliz tilidagi korpuslarning eng so‘nggi avlodi sifatida *Ingliz tili banki (The Cobuild Project/ The Bank of English), Britaniya milliy korpusi, Xalqaro ingliz tili korpusi (The International Corpus of English (ICE)), Amerika milliy korpusi (American National Corpus)* kabi mashhur korpuslarning mavjudligi Milliy til va Davlat tili taraqqiyotida Milliy korpusning ahamiyati va o‘rnining qanchalik muhimligini ko‘rsatadi.[8]

Milliy til taraqqiyotida axborot-kommunikatsion texnologiyalar, kompyuter lingvistikasining o‘rni beqiyos. O‘zbek tili tarixida kompyuter lingvistikasi rivoji matematika, amaliy matematika, informatika fanlarining keng doiradagi rivoji va elektron hisoblash mashinalarining taraqqiyoti bilan bog‘liq. *Professor Baxtiyor Mengliyevning* Milliy korpus haqida bildirgan quyidagi fikrlari juda asosli va o‘rinli edi: “**Milliy til korpusi – milliy til birliklarining xususiyatlarini aniqlash maqsadida elektron qidiruv imkoniyati mavjud tizim, tabiiy tilning raqamlashgan yozma va og‘zaki matnlari jamlanmasi.** Milliy korpus – tilni

[7] Hamroeva Sh. O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari. Filol.fan.bo‘yicha falsafa doktori (PhD)…dis. aftoref. – Qarshi, 2018.
[8] Ҳамроева Ш. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент, 2023.

tadqiq etish (til birligining o‘zgarishi, eskirishi, yangilarining paydo bo‘lishi, ma’nosining kengayishi va torayishi; yangi iboralarning paydo bo‘lishini kuzatish), o‘rganish, an’anaviy va zamonaviy lug‘atlar tuzishda keng imkoniyatli dasturlashtirilgan tizim. Shu bois Milliy korpus milliy til xazinasi demakdir. Korpusdagi milliy so‘zi nafaqat tilning, balki korpus tuzilishi va tarkibining ham o‘ziga xosligini anglatadi. Milliy korpus boshqa tur til korpuslaridan farqi majmuaviylik tabiati – unda ta’limiy korpus, mualliflik korpusi, poetik matnlar korpusi, ilmiy va rasmiy matnlar korpusi, og‘zaki matnlar, badiiy matnlar va gazetalar hamda dialektlar korpusi kabi ichki korpuslarni o‘z tarkibiga qamrab olganligi bilan belgilanadi. Shu bois milliy korpusni yaratishda katta jamoa faoliyati zarur.

Milliy korpusdan lingvistlar, leksikograflar, kompyuter lingvistlari, dasturchilar, muharrirlar, tarjimonlar, jurnalistlar, noshirlar, olimlar, o‘qituvchilar, ta’lim oluvchilar va boshqa har qanday soha mutaxassisi keng foydalanish imkoniyatiga ega bo‘ladi. Asrimizda korpussiz na nazariy, na amaliy filologiya taraqqiy etadi”.[9]

Til korpuslari – til bo‘yicha tadqiqot va amaliy topshiriqlar yechimi uchun zarur ish quroli. O‘zbek tilidagi korpuslar shu tilda yaratilgan korpuslarning barcha turini o‘zida jamlaydi. Natijada muayyan tilda shu tilning barcha imkoniyatlarini o‘zida aks ettirgan ma’lumotlar bazasi shakllanadi. O‘zbek tili milliy korpusi yaratilsa, tilni kompyuter texnologiyalari yordamida o‘rganuvchi, ta’lim oluvchi uchun bir qancha afzalliklarni yaratadi va o‘zbek tilidagi ma’lumotlarni saqlash, qayta ishlash imkonini beradi. Shuning bilan birga, Milliy korpus dunyoning ko‘plab nuqtalaridagi o‘zbek tilida muloqot qiluvchi barcha insonlar uchun foydali va muhim ekanligini ham alohida ta’kidlash lozimdir. Sababi, bugungi kunda mavjud bo‘lgan korpuslar til o‘rganish uchun ham qulay, ham juda zarur vositadir. O‘zbek tili Milliy korpusini yaratish, takomillashtirish

[9] Mengliyev B., Bobojonov S., Hamroyeva Sh. O‘zbek tilining milliy korpusi. – Toshkent: Ma’rifat//2018.

yo‘lidagi bu kabi harakatlarni davrning o‘zi taqozo etmoqda. Dunyo kompyuter lingvistikasida Milliy til korpusining mavjudligiga tilning yashovchanlik va kompyuter tiliga aylanish mezoni sifatida qaralmoqda. Biz ham bugungi kun talablaridan kelib chiqib o‘zbek tilining milliy korpusini tuzishni kechiktirib bo‘lmas vazifa deb hisoblashimiz va amaliy harakatlarni jadallashtirishimiz kerak. Milliy til korpusi – ona tilimizning jahonga yuz tutishiga erishish, milliy ma’naviyatimizni takomillashtirish va yuksaltirishning ham asosiy ham yagona vositasidir.

Foydalanilgan adabiyotlar ro‘yxati:

1. Mengliyev B., Bobojonov S., Hamroyeva Sh. O‘zbek tilining milliy korpusi. – Toshkent: Ma’rifat//2018.
2. Менглиев Б., Ҳамроева Ш. Лингвистик таъминот воситаларининг умумий тавсифи ҳамда тил корпусини яратишдаги аҳамияти. – Тошкент, 2019.
3. Po‘latov A. Kompyuter lingvistikasi. – Samarqand, 2011.
4. Rahimov A. Kompyuter lingvistikasi asoslari. – Toshkent: Akademnashr, 2011.
5. Raxmanova A. Korpus lingvistikasiga oid terminlar lug‘ati. – Toshkent, 2021.
6. Zaxarov V., Mengliyev B., Hamroyeva Sh. Korpus lingvistikasi: o‘quv qo‘llanma. – Toshkent, 2023.
7. Ҳамроева Ш. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент, 2018.