



PARALLEL TIL KORPUSLARINING TIL O'RGANISHDAGI O'RNI

Nurullayeva Oydin
TDO 'TAU doktoranti

Annotatsiya: Maqolada korpus lingvistikasi, korpus va uning parallel korpus haqidagi fikrlar, uning tuzilishi, korpus turlari, til korpuslarning avtomatik tarjimadagi o'rni haqida so'z boradi. Bugungi kunda korpusning nazariy hamda amaliy ahamiyati o'zbek tilshunosligida tilning mavjud imkoniyatlarini kengroq o'rganishda, elektron lug'atlar yaratishda, tilni o'rganishda zamonaviy axborot-texnologiyalaridan foydalanish samaradorligini oshirishda, tilda avtomatik tarjima, qidiruv va kompyuter tahlili kabi masalalarni hal qilishda, tilning maxsus sohalar bo'yicha korpusini qurish zarurati mavjudligini ko'rsatadi.

Kalit so'zlar: Korpus, korpus lingvistikasi, parallel korpus, Comparable/Multilingual Corpora, OPUS korpusi, avtomatik tarjima.

Mamlakatimizda oxirgi yillarda barcha sohalarda ish samaradorligini oshirish, insonlar mehnatini yengillashtirish maqsadida kompyuter texnologiyalaridan foydalanish bo'yicha zarur islohotlar olib borilmoqda. Jumladan, o'zbek tilining maqomini oshirish hamda imkoniyatlarini kengaytirish uchun bugungi kunda korpus lingvistikasi hamda uning sohalari jadallik bilan rivojlanmoqda.

D.Dobrovolskiy korpus lingvistikasining muhim yo'nalishi bo'lgan parallel korpusga shunday ta'rif beradi: "Parallel korpus – asl va tarjimadagi elektron matnlar yig'indisidan iborat korpus. Asl va tarjima matnlar shunchaki yonma-yon qo'yilmaydi, balki bu ikki matndagi gap(sistaktik birlik)lar mazmunan mos kelish darajasida bo'lakka bo'linib, mana shu birliklar yonma-yon, bir-biriga ishora/havola qilish holatida turadi. Asliyatdagi matn fragmentiga mos keluvchi tarjimadagi fragment belgilab qo'yilgan bo'ladi. Aynan mana shu holat bunday korpuslar yordamida turli lingvistik amallar bajarishga imkon yaratadi".

Zamonaviy korpus lingvistikasida parallel korpusning ikki ko'rinishi mavjud:

- 1) ko'p tilli korpus (Comparable/Multilingual Corpora);
- 2) tarjima korpus (Translation Corpora).

Bunday xususiyatga ega korpusning struktur tarkibi uning maqsadidan kelib chiqib turlicha bo'lishi mumkin: 1) tarjimaga havola qiluvchi odatiy matn tarzida; 2) qiyoslash uchun qulay bo'lgan «oynadagi matnlar» shaklida; 3) ma'lumotlar bazasi ko'rinishida.



Parallel korpuslar tarjimashunoslik va qarama-qarshi tilshunoslikda markaziy o'rin tutadi. Ko'pgina parallel korpuslarga foydalanish oson konkordanserlar orqali kirish mumkin, bu esa tillararo hodisalarni o'rganishni sezilarli darajada osonlashtiradi. Anarbayerev Orzubek Raxmanovich parallel korpusga quyidagicha fikr beradi: „ ... Parallel korpus madaniyatlararo muloqot keng yoyilgan hozirgi davr uchun muhim voqelik hisoblanadi. Parallel korpuslar orqali turli til muhitidagi, madaniyatlaridagi universalialar hamda tillarning o'ziga xos mental xususiyatlari, realiya va lakunar birliklarni aniqlash mumkin bo'ladi. Parallel matnlar korpusi avtomatik tarjima rivoji uchun ham xizmat qiladi, kompyuter leksikografiyasining taraqqiyotini ta'minlaydi. Parallel matnlar korpusi yordamida konkordanser dasturlar ishlab chiqiladi va turli xil ixtisoslik lug'atlari tuzish imkoniyati yuzaga keladi “Darhaqiqat, korpuslar yaratilishi, eng avvalo, leksikografiya sohasi uchun katta ahamiyatga ega. Zamon aksi tilning lug'atida namoyon bo'ladi. Doimiy ravishda “o'sib” boradi. Tilning qay darajada boyligini yaqqol ko'rsatadigan oynasi – bu uning lug'ati. Lug'at tarkibining ko'pligi uning qadrini oshiradigan omil sanaladi. Bunday korpuslar ham til o'rgatish uchun boy materiallar manbai hisoblanadi.

Bundan tashqari, parallel korpuslar statistik mashina tarjimasi tizimlari uchun o'quv ma'lumotlari sifatida xizmat qiladi. Parallel korpus bu – tarjima qilingan matnlar juftligidir. CLARIN kabi parallel korpuslar bizning eng katta resurs bazamizdir, infratuzilma 87 ta parallel korpusga kirishni ta'minlaydi. Tarjimashunoslikda asosiy e'tibor tarjimalarni asl matnlardan ajratib turadigan xususiyatlarni aniqlashga qaratilgan. Ushbu o'zgartirishlar ma'lum bir tarjima vazifasi yoki tarjima juftligi uchun individual bo'lishi mumkin, lekin ular tarjima matnining keng ko'lamli lingvistik xususiyatlariga ko'ra tarjimalarni tarjima qilinmagan matnlardan farq qiladigan odatiy xususiyatlarni aniqlashtirishi mumkin.

Ma'lumki, bugungi kunda har bir sohada parallel korpuslar turli amaliy maqsadlarda qo'llaniladi. Parallel korpuslar ikki tilda lingvistik xususiyatlar va ularning chastotalarini solishtirish uchun ishlatiladi. Ular, shuningdek, manba va maqsadli tillar o'rtasidagi o'xshashlik va farqlarni o'rganish uchun ishlatiladi. Bu tahlilning turli darajalarida tizimli, matnga asoslangan kontrastli tadqiqotlar o'tkazish imkonini beradi. Shu tarzda, parallel korpuslar tillar haqida lingvistik tipologik va madaniy farqlar va o'xshashliklarga asoslangan ma'lumot beradi. Qiyosiy tilshunoslik parallel korpuslardan foydalanish bilan chambarchas bog'liq, ya'ni ular tarjimashunoslikda ikki yoki undan ortiq tillarni qiyosiy tahlil qilish imkonini beradi. Parallel korpus tarjimonlarga asl manba va maqsadli til o'rtasidagi tarjima ekvivalentlarini topishga ham yordam beradi.



Liverpul universitetida 2009-yilgi Korpus tilshunosligi konferensiyasi bo'lib o'tadi, unda parallel korpuslar bilan ishlaydigan tilshunoslar va tarjimashunoslik talablari, korpusdan o'z maqsadlari uchun foydalanish asboblari hamda korpus interfeysi bilan bog'liq masalalar muhokama qilinadi. Ushbu konferensiya Lankaster universitetida 2003-yilda Korpus tilshunosligi konferensiyasida o'tkazilgan "Multilingual Corpora: lingvistik talablar va texnik istiqbollar" seminari bilan chambarchas bog'liq va uning davomi edi. Muhokamada Volk, sintaktik darajada annotatsiya, moslashtirish va so'rovlarga e'tibor qaratish kerakligini ta'kidlaydi. Fantinuoli esa konferensiya tarjimonlariga ishining barcha bosqichlarida yordam beradigan InterpretBank dasturiy ta'minotining dizaynini tasvirlaydi. U Boot Cat mexanizmiga asoslanib, qidiruv so'zlari to'plamini berilgan domenga xos hujjatlar uchun Internetdan yig'adi, ular bo'yicha atamalarni ajratib oladi va qo'shimcha resurslardan foydalanadi. Vikipediya, ikki tilli onlayn lug'atlar, ta'riflar, tarjimalar, birikmalar va kontekstdagi kalit so'zlar haqida ma'lumot chaqiradi. Barcha modullar tarjimonlarning o'ziga xos ehtiyojlariga moslashtirilgan bo'lib, tayyorlash uchun zarur bo'lgan vaqtni qisqartiradi va tarjimon paytida samarali qidirish imkonini beradi.

OPUS korpusi tayyorgarlik bosqichida ekanini aytadi. Parallel korpusda tarjima o'zgarishlarini aniqlash tarjimashunoslik nuqtai nazaridan asosiy vazifadir. Tarjimashunoslikda ham, qarama-qarshi tilshunoslikda ham ko'p tilli korpuslar yaqinda tarjima hodisalarini, ya'ni tarjima o'zgarishlarini yoki tarjima xususiyatlarini o'rganish uchun, shuningdek, o'rtasidagi kontrastiv farqlarni o'rganish uchun foydalanilgan. Shunday korpuslardan biri ingliz-german CroCo korpusidir. Korpus tarkibiga ingliz va nemis asl nusxalari va ularning nemis va ingliz tillariga tarjimalari kiradi. Shunday qilib, u ham taqqoslanadigan, ham parallel korpus sifatida ishlatilishi mumkin. Har bir tarjima yo'nalishlari har bir bo'limda jami 31250 so'zdan iborat bo'lgan kamida 10 ta matndan iborat 8 ta janrda aks ettirilgan. CroCo Corpus taxminan bir million so'zni o'z ichiga oladi. Bundan tashqari, ro'yxatdan o'tmagan neytral ma'lumot korpusi 17 ta janrdan 2000 ta so'z namunalarini o'z ichiga olgan nemis va ingliz tillari uchun kiritilgan. Shunday qilib, korpus o'xshash va parallel qismdan iborat. Janrlar jihatidan siyosiy insholar, badiiy matnlar (badiiy adabiyot), o'quv qo'llanmalari, ilmiy-ommabop matnlar, korporativ muloqot, tayyorlangan nutqlar, turizm varaqalari va veb-saytlar (WEB)ni o'z ichiga oladi. Unda tobora murakkab so'rovlarni, shu jumladan bir nechta izoh va moslashtirish qatlamlarida ishlaydigan so'rovlarni boshqarish va Java-ga asoslangan annotatsiya vositalarini korpusga qo'llash uchun o'z amaliy dasturlash interfeysi (API) yaratilgan. Ular ingliz-



nemis tillari juftligida tarjima xususiyatlarini o'rganish uchun dolzarbligi sababli tanlangan.

Parallel korpuslarning yana bir zarur tomoni – tarjimadagi g'alizliklarga chek qo'yishidir. Internet orqali qilingan tarjimada ko'plab xatoliklar, so'zlarning umuman matn tarjimasiga aloqasi bo'lmagan so'z sifatida o'girib berishi haligacha muammo. Parallel korpus yaratishdan asosiy ko'zlanayotgan maqsadimiz ham samarali, to'g'ri tarjima usullarini ishlab chiqishdir.

Foydalanilgan adabiyotlar ro'yxati:

1. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
2. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (JFLTAL). 2019;6(1-2019):131-7.
3. Захаров В. П., Богданова С. Ю. Корпусная лингвистика: учебник. 3-е изд., перераб. — СПб.: Изд-во С.-Петербур. ун-та, 2020. — 234 с.
4. Михайлов М. Параллельные корпуса художественных текстов: диссертация.- Финляндия . ун-та, 2003. —348с.