

KORPUS LINGVISTIKASIDA TARJIMASHUNOSLIK MASALASI

Abdurahmanova Muqaddas Tursunaliyevna

MParfi2005@yandex.ru

O'zbekiston Milliy universiteti Jurnalistika fakulteti

O'zbek tilshunosligi kafedrasida dotsenti, f.f.n.

Rayimjonova Matluba Maxamadkarimovna

rayimjonova89@mail.ru

O'zbekiston Milliy universiteti Jurnalistika fakulteti

Kompyuter lingvistikasi yo'nalishi 1-kurs magistranti

Annotatsiya: Ushbu maqolada parallel korpusning matn tarjimasida o'ziga xos o'rni hamda tillarni o'rganish va bir-biriga qiyoslashdagi ahamiyati yoritib berilgan. Shuningdek, parallel korpus tuzish texnologiyasiga ham alohida e'tibor qaratilgan.

Kalit so'zlar: *mashina tarjimasi, korpus, korpus lingvistikasi, parallel korpus, parallel konkordans, bitex.*

Annotation: This article will discuss the role of the parallel corpus in text translation and its importance in the study and comparison of languages. Moreover, the technology compiling of parallel corpus was taken into special account.

Keywords: machine translation, corpus, corpus linguistics, parallel corpus, parallel concordans, bitex.

Milliy korpus milliy til xazinasidir demakdir.

B.Mengliyev

Korpus lingvistikasining taraqqiyot tamoyillarini, korpus yaratishning kompyuter usullari, matematik modellarini belgilash, parallel korpuslar, kompyuter lug'atlarining ma'lumotlar bazasi sifatidagi ahamiyatini yoritish, tezaurus va konkordanslarning lingvistik ta'minotini izohlash, korpus turlarini tahlil qilish, milliy til taraqqiyotidagi o'rnini ko'rsatish, ijtimoiy sohalar rivojida, ta'lim jarayonidagi samaradorligini aniqlash muhim qiymatga ega bo'lmoqda. So'nggi yillarda amaliy tilshunoslik sohalarini rivojlantirish, kompyuter lingvistikasi yo'nalishlarini, lingvistik masalalarni kompyuter imkoniyatlari asosida hal qilish, o'zbek tilining davlat tili sifatidagi funksional doirasini kengaytirish masalalariga alohida e'tibor berila boshlansada, bu borada sanoqli ishlar amalga oshirildi,



tizimli faoliyat olib borilmadi. Kirill alifbosidan yangi lotin alifbosiga o'tish, alifboni o'zgartirish, yangi imlo qoidalarini o'zlashtirish, o'zbek tili grammatikasini rus tili qolipidan farqlash bilan bog'liq vazifalar amalga oshirildi. Bu holat esa o'zbek tilini kompyuter tiliga aylantirish, o'zbek tili milliy korpusini shakllantirish, kompyuter va tarjima lug'atlarini yaratish jarayonini tezlashtirishga monelik qildi. Amaliy tilshunoslikni rivojlantirish, o'zbek tilini internet tiliga aylantirish, til masalalarini kompyuter yordamida hal qilish muammoligicha qolaverdi. Respublikamizda davlat tiliga e'tibor milliy qadriyatlarni e'zozlashning ustuvor yo'nalishlaridan biri darajasiga ko'tarildi. Milliy tilni asrab-avaylash, undan amaliy foydalanish ko'nikma hamda malakalarini shakllantirish, shu bilan bir qatorda, o'zbek tilining zamonaviy axborot-kommunikatsiya tizimida keng qo'llanishini ta'minlash dolzarb vazifaga aylandi. “Davlat tilining axborot va kommunikatsiya texnologiyalari, xususan, Internet jahon axborot tarmog'ida munosib o'rin egallashini ta'minlash, o'zbek tilining kompyuter dasturlarini yaratish” muhimligi ta'kidlandi [O'zRP Farmoni, 2019. № 5850].

Amaliy lingvistikaning alohida sohasi hisoblangan Korpus lingvistikasi o'qitish metodikasini takomillashtirishga asoslangani, grammatik xususiyatlarni tushuntirishda ko'nikma hamda malakaga e'tibor qaratilgani, chet tillarini, milliy tilni ona tili va chet tili sifatida o'qitish masalalari, avtomatik tarjima bilan shug'ullangan uchun unga bo'lgan qiziqish ortib bormoqda.

Korpus lingvistikasi milliy til taraqqiyotida o'ziga xos qiymatga ega. Xususan, avtomatik mashina tarjimasi uchun ahamiyatlidir. Ma'lumki, tarjima – bir tildagi matnni boshqa tilda qayta yaratishdan iborat adabiy ijod turi. Tarjima amaliyotini, uning o'ziga xosliklarini, tarixi, tamoyillari va qonuniyatlarini tarjimashunoslik fani o'rganadi, tarjimonga o'girish uchun asar tanlash, tillararo tafovutlar hamda asliyatga xos boshqa xususiyatlarni, milliy adabiy an'analarni hisobga olgan holda ish tutishda yordam beradi. Tarjimashunoslik XX asrda rivojlandi. Yevropada tarjimashunoslik bo'yicha jamiyatlar tuzildi. O'zbekistonda tarjimashunoslik maxsus ilm sifatida o'tgan asrning 2-yarmidan shakllandi. Asr oxiriga kelib, bu sohada 10 ga yaqin fan doktori, 50 dan ortiq fan nomzodi yetishib chiqdi. Shu bilan bir qatorda AQShda 1950-yillarda chet el tillaridan, xususan, rus ilmiy jurnallaridan matnlarni avtomatik ravishda ingliz tiliga tarjima qilish uchun kompyuterlardan foydalanish harakatlari bilan mashina tarjimasi ham vujudga kela boshladi. Mashina tarjimasi tizimida qaysi tildan qaysi tilga tarjima qilinsa, o'sha tillarning ham grammatik, ham semantik, imloviy xususiyatlarini aks ettiradigan maxsus dastur tuzib olish kerak bo'ladi. Zarur hollarda mashina tarjimasi ishiga

inson-muharrir ham aralashadi. Bugungi kunda tarjima nazariyasi va amaliyoti, inson amaliy tarjimasidan tortib mashina tarjimasiga va terminologiyasiga qadar korpus tilshunosligi asosiy paradigma va tadqiqot metodologiyasiga aylandi [Claudio Fantinuoli, 2015]. Korpusning mashina tarjimasiga uchun foydali jihatlari shundaki, tarjima qilish jarayoni amalga oshirilganda dastlabki ish lug'atdan so'z qidirishdan boshlanadi, ya'ni EHM tarjima qilinadigan matndan so'zlarni birin-ketin olib, ularning har biriga mos keladigan asosni lug'atdan qidirib topadi. Keyin ularni taqqoslaydi, harflar soni jihatidan maksimal darajada mos keluvchi so'zga sig'adigan asosni qidirib topadi va uni tarjima qiladi. Korpus mashina tarjimasiga uchun katta leksikografik baza hisoblanadi [Rahimov, 2011]. Korpus lingvistik tarjima ekvivalentlarini avtomatik tahlil qilish va rasmiylashtirish bosqichi uchun asos bo'lib xizmat qiladi.

Ona tili va chet tillarini o'qitish va o'rganish borasida korpusning o'rnini beqiyos. Izlanuvchi korpusga kiruvchi materialni mustaqil ravishda o'rganib, u yoki bu til birligining mohiyati, nutq tuzishdagi ahamiyati, imkoniyatlari haqida taxmin, xulosalar qilishi mumkin. Mazkur induktiv metod yordamida o'zlashtirilgan bilim uzoq muddat uning xotirasida saqlanishi barchaga ma'lum. Masalan, ikkinchi tilni o'rganayotgan talaba bir-biriga o'xshash ikki so'zning farqini tushuntirib berishi kerak, shu o'rinda u mavjud korpus materiallaridan foydalanib, yuqoridagi so'zlar qatnashgan 20-30 ta gapni bir zumda topadi hamda kontekst tahlil jarayonida o'xshash so'zlarning semantik farqini tushunib yetadi. Bugungi kunda kompyuterlashtirilgan ta'lim o'quv materialini taqdim etish usulidan biroz chekinib, ijodkorlikni talab qiluvchi metodlarga yuzlanmoqda [Carol A., 2003]. Aynan til o'rgatishni maqsad qiluvchi korpuslar ham mavjud bo'lib, «Учебный корпус русского языка», «Learner corpus of English» shular jumlasidandir. Xorijiy til vakillari bilan ishlash jarayonida til korpusining ahamiyati bir necha marta ortadi.

Korpus lingvistikasida parallel matnlar korpusi muhim ahamiyat kasb etadi. Parallel korpuslar bu tarjimani o'qitishga va chet tilini o'rgatishga qaratilgan korpuslardir.

Parallel korpuslarda tarjima masalasi bugungi kunda dolzarb masalalardan biri hisoblanadi. Korpus mutaxassislarini (asosan, tarjimon) doim bir necha tilli korpus yaratish qiziqtirib kelgan. Korpus yaratishning ilk davridan boshlab ingliz, fin, fransuz, nemis, grek, norveg, ispan, shved va hokazo tillar uchun ikki tilli korpuslar paydo bo'la boshlagan. Bunday korpus bitex deb ham ataladi. Korpusni ikki tilli emas, balki uch, to'rt va undan ortiq tilli qilishga hech qanday to'siq yo'q.

Mutaxassislar parallellik nuqtayi nazaridan korpusni bir, ikki va ko'p tilli kabi turlarga ham bo'lishadi. Bir tilli korpusda til varianti va shevalar bir-biriga qarama-qarshi qo'yilsa, ikki va ko'p tilli korpus bir mavzu doirasida turli tilda yozilgan matnlar majmuidan iborat bo'ladi. Masalan, ma'lum ilmiy muammo borasida turli davlatda turli tilda o'tkazilgan konferensiya materiallarini qamrab olishi mumkin. Ko'p tilli korpuslar, odatda, tarjimonlar tomonidan foydalaniladi. Ko'p tilli korpusning yana bir ko'rinishi original matn va tarjima matndan iborat bo'ladi. Korpusning ushbu turi qiyosiy chog'ishtirma tadqiqot olib borishda, tarjima nazariyasi hamda kompyuter tarjimasini o'rganishda juda muhim manba bo'lib xizmat qiladi.

Ko'p tilli korpusning 2 turi mavjud:

- 1) bir-birining tarjimasi bo'lgan matnli korpus;
- 2) bir mavzuga oid ikki tildagi matnli korpus.

Birinchi tipdagi korpus —parallel korpus (parallel corpora) deb nomlanib, ma'lum bir tarjimaning turli aspektini o'rganish uchun qo'llaniladi.

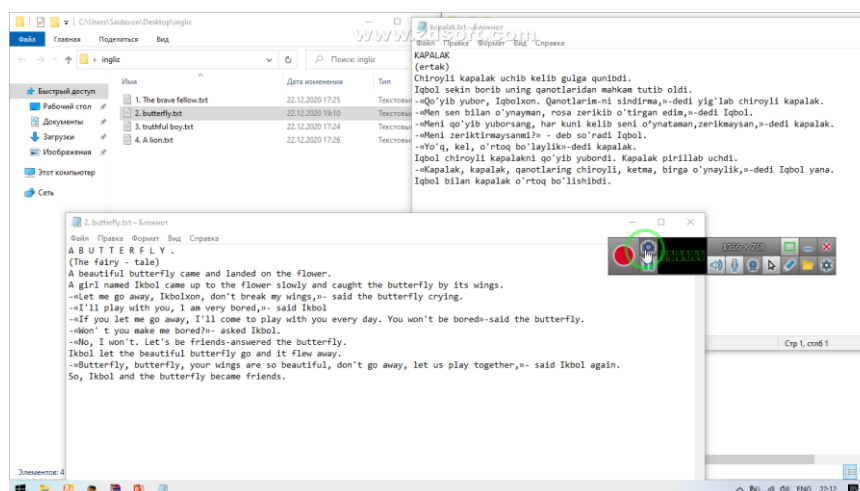
Parallel korpusning qiymati uning hajmi va tillarning miqdori bilan belgilanadi. “Acquis Communautaire” dunyodagi eng katta parallel korpus bo'lib, muhim jihati bu korpusdan foydalanishning bepulligi va multi-eston, sloven-fin kabi kam uchraydigan tillar juftligining mavjudligi bilan baholanadi. Ushbu korpuslardan quyidagi maqsadlarda foydalanish mumkin:

1. Tipik tarjima usullari va transformatsiyani yuzaga keltirish;
2. Avtomatik tarjima tizimi statistikasini o'rganish;
3. Bir va ko'p tilli lug'atlar yaratish;
4. Ma'lumotni saqlash va uzatish dasturlarini o'rganish va baholash;
5. Tarjima to'g'riligini avtomatik tarzda tekshirish;
6. Ekvivalent tanlash imkoni kengligi orqali tarjimon mehnatini osonlashtirish.

Parallel korpus yaratish muhimligini inobatga olib parallel korpus yaratish texnologiyasini o'zbek va ingliz tilidagi matnlarda qo'llab ko'rdik.

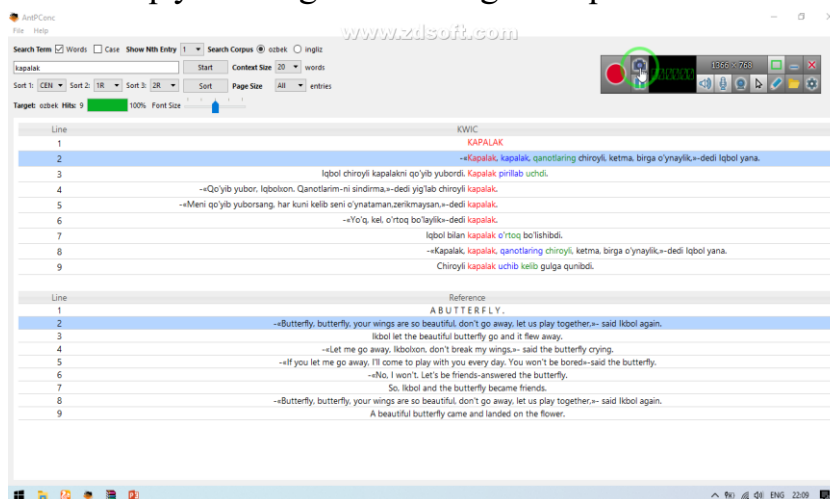
1. Parallel korpus uchun olingan matnning ikki bob yoki kamida ikkita xatboshi bir-biriga to'g'ri keladigan bo'lsa, bu matn tarjima munosabatlariga ega bo'ladi.

2. Agar kamida ikkita jumla bir-biriga to'g'ri keladigan bo'lsa, ikkita xatboshida tarjima aloqasi mavjud bo'ladi.



3. Agar kamida ikkita soʻz bir-biriga toʻgʻri keladigan boʻlsa, ikkita jumla tarjima munosabatlariga ega.

4. Ikkala soʻzning tarjima aloqasi, agar ularning kamida bitta maʼnosi mos yozuvlar sifatida ishlatilgan ikki tilli lugʻat tomonidan tasdiqlangan boʻlsa bir-birini taqozo qiladi. Shuni ham taʼkidlash kerakki, raqamlar (sana va raqamlar) ikki tilda aniq yoʻnaltiruvchi nuqtalar va ichki jumla birliklarining ishonchli identifikatorlari boʻlishi kerak. Muammoni keltirib chiqaradigan hodisalardan biri-mavqeyi ikki tilda bir xil boʻlmagan iboralaridir. Biroq, ehtimollik hisob-kitoblardan foydalanib, yoʻq qilish jarayoni bilan, qolgan birliklarning tengligi boshqa barcha birliklarni qayta ishlagandan soʻng tasdiqlanishi mumkin.



Korpus lingvistikasi Amaliy tilshunoslikning alohida rivojlantirilishi kerak boʻlgan sohalaridan biridir. Milliy korpuslarning har bir koʻrinishi milliy til taraqqiyotiga xizmat qiladi. Oʻzbek tili va qardosh boʻlmagan tillar korpusi madaniyatlar oʻrtasidagi universal hamda farqli jihatlarni namoyon qilishga xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR RO‘YXATI:

1. O‘zbekiston Respublikasi Prezidentining “O‘zbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan oshirish chora-tadbirlari to‘g‘risida”gi PF-5850-son Farmoni // Xalq so‘zi, 2019-yil 22-oktyabr. № 218 (7448).
2. Захаров В.П. Богданова С. Корпусная лингвистика. Учебное пособие. – Санкт-Петербург, 2020
3. <https://postnauka.ru/video/7783>. Плунгян В. Корпусная лингвистика // ПостНаука [Электронный ресурс]. 2013.
4. Claudio Fantinuoli & Federico Zanettin. 2015. Creating and using multilingual corpora in translation studies. In Claudio Fantinuoli & Federico Zanettin (eds.), *New directions in corpus-based translation studies*, 1–11. Berlin: Language Science Press
5. Azamat Rahimov. Kompyuter lingvistikasi asoslari. O‘quv qo‘llanma. – Toshkent, 2011.
6. Carol A. Chapelle // *Essential teacher*. 2003. Vol. 9. P. 5-11. - P. 6.
7. Baker, M (1993). "Corpus linguistics and translation studies. "Implications and applications" in M. Baker G. Francis & E. Tognini-Bonelli (eds.) *Text and technology*. Philadelphia/ Amsterdam: John Benjamins, 232–252.
8. Scott, M. (1996) *Wordsmith tools* .Oxford: Oxford University Press.
9. <https://uz.wikipedia.org/wiki>
10. <https://laurenceanthony.software.html>