

## PARALLEL KORPUS XUSUSIYATI, TUZILISHI, INTERFEYSI

**Nazira G`anijon qizi Sobirova**

NamMQI magistranti

**Madina Akramajanova**

O`zMU magistranti

**Shohjahon G`anijon o`g`li Sobirov**

NamMQI talabasi

### ANNOTATSIYA

Ushbu maqolada korpus lingvistikasi, korpus va uning parallel korpus bog`ini haqidagi fikrlar, uning tuzilishi, korpus turlari, token, lemma, stemming haqida so`z boradi. Bugungi kunda korpusning nazariy hamda amaliy ahamiyati o`zbek tilshunosligida tilning mavjud imkoniyatlarini kengroq o`rganishda, tilshunoslikdagi muammoli tomonlarini aniqlash, 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, tarjima korpus, taqqoslanadigan korpus, biteks, tekislash

### KIRISH

Parallel korpus (Parallel Corpora) – parallel tarjima matnlarining elektron analogi; ko`plab «original matn va ularning bir/bir necha tarjimasi» bloklaridan iborat. Korpusdagi elektron matnlar original matnning o`zi yoki uning bir qismi bo`lishi mumkin. Parallel korpus – original matn va uning tarjimasi yig`ilgan korpus.

Parallel korpusning tuzilishi, tarkibi va imkoniyati haqidagi ma`lumotlarni D.O.Dobrovolskiy, Yu.Tao, V.Zaxarov, A.A.Kokoreva, E.P.Sosninalarning ishlarida kuzatamiz.

### ADABIYOTLAR TAHLILI VA METODOLOGIYA

D.Dobrovolskiy 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 tili 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.

Tuzilishi va matnlarning joylashuvi, birliklarining moslashtirilishiga ko'ra parallel korpusning bir necha ko'rinishi mavjud:

- 1) bir yo'nalishli (ingliz tilidan rus tiliga tarjima qilingan matn)
- 2) ikki yo'nalishli (ingliz tilidan rus tiliga tarjima qilingan matn va rus tilidan ingliz tiliga (teskari tarjima))
- 3) ko'p yo'nalishli (ingliz tilidan rus, nemis, fransuz tillariga tarjima qilingan matn)

Parallel korpus tarjima tilining struktur tuzilishi, ibora va so'zlarning ma'lum kontekstdagi imkoniyatlarini kuzatishga xizmat qiladi. Bunday korpuslar quyidagi maqsadlarda foydalaniladi:

- 1) qiyosiy tilshunoslikda (ikki tilstrukturasini tahlil qilish maqsadida);
- 2) tarjima sohasida (original matn ekvivalentini boshqa tillardan qidirish maqsadida);
- 3) avtomatik tarjima sohasida;
- 4) lingvodidaktikada;
- 5) leksikografiyada.

Matn va uning tarjimasi yonma-yon keladigan elektron to'planning boshqa ko'rinishlari ham mavjud. Bitekst dasturi shunday vositalar sirasiga kiradi. Parallel matn (bitekst)– boshqa tildagi tarjimasi bilan beriladigan matnlar to'plami.

Bitekstlar «moslashtiruvchi asbob» (*alignment tool*) yoki «bitekst uchun moslama» (*bitext tool*) deb nomlanuvchi maxsus kompyuter dasturlari orqali amalga oshiriladi. Bu dasturlar original va tarjima matn mazmunini turli sintaktik birliklar, asosan, sodda

gap shaklida moslashtirib beradi. Bitekstar yig'indisi *bitekst ma'lumotlar bazasi* yoki *ikki tilli korpus* deb nomlanadi ham turli bog'lanishlarni kazatish imkonini beruvchi ma'lumot bazasi (spravochnik) vazifasini bajaradi.

*Parallel korpus: ilk va keyingi ko'rinish.* Korpus lingvistikasi yo'nalishi xorijiy tillarni o'qitish metodikasi va tarjima, kompyuter lingvistikasi yo'nalishi uchun amaliy ahamiyat kasb etmoqda. Xorijiy til va tarjima nazariyasi ta'limida parallel korpuslarning ahamiyati borasida M.Barlou, Makkeneri, Beyker, Zanneti, Arenberga, Blank, Braun, Cherch, Geyl, Devis, Foster kabi olimlarning fikrlari qimmatli.

Bitekst g'oyasi Brayan Xerrisga tegishli bo'lib, u dastlab 1988-yilda bitekst konsepsiyasini ishlab chiqqan; ushbu konsepsiya keyinchalik Monreal univesiteti bir guruh olimlari tomonidan rivojlantirilgan; RALI (*Recherche appliquée en linguistique informatique* yoki *Applied Research in Computational Linguistics// Прикладные исследования в вычислительной лингвистике*) deb nomlanuvchi uyushmada faoliyat olib borilgan. Guruhda tilshunos va dasturchilar faoliyat yuritgan; matnni qayta ishlash muammolarini o'rganishgan. Ilk marta bitekst konsepsiyasini Pyer Izabel va Klod Bedarlar amalga oshirishgan.

“Parallel matnlar korpusi ilmiy va amaliy maqsadlarda (shu jumladan, xorijiy tillarni o'qitish maqsadida) amalga oshirilmoqda. Parallel matnlar strukturasi ko'ra manba tildan maqsad tilga tarjima matnlardan iborat. Masalan, «Alie in Wonderland» ingliz matni va uning nemis, fransuz, rus tillaridagi tarjimasi parallel matnlarni yaratishga asos bo'ladi”.

## MUHOKAMA VA NATIJALAR

1897-98-yillarda nemis tilshunosi Keding tomonidan birinchi korpus yaratilgan; u qog'oz shaklda bo'lib, so'zlarda tovushlarning taqsimlanish chastotasini tekshirish maqsadida tuzilgan edi. Ammo katta hajmli materialni qo'lda hisoblab chiqish juda mushkul mashg'ulot ekanligini hozirgi tadqiqotchilar biladi. Keyinchalik paydo bo'lgan konkordans yoki undan ham mukammalrog'i - korpus menejeri bu vazifani ancha yengillashtirdi. Korpus tuzish tajribasi shuni ko'rsatdiki, universal korpus yaratishning iloji yo'q. Tadqiqotchi maqsadiga ko'ra korpus imkoniyatlari chegaralanadi va har bir vazifani alohida korpuslar bajaradi.

Korpusga qo'yiladigan talablardan biri uning interfeysga egaligidir.

Korpus lingvistikasida o'zbek-turk parallel matnlar korpusi muhim ahamiyat kasb etadi. O'zbek-turk parallel matnlar korpusi o'zbek hamda turk tillaridagi badiiy asar, qo'llanma, ommaviy axborot vositalari, turli xil hujjatlarning ikki yoki undan ko'p tillaridagi

elektron ko‘rinishlaridir. Bunday korpuslar turkologiyaga oid tadqiqotlar uchun ahamiyatlidir.

O‘zbek-turk parallel matnlar korpusi izohli o‘quv qo‘llanma, o‘quv adabiyotlari matni bo‘yicha tuzilgan korpuslar hisoblanib, tillar o‘rtasidagi umumiy va farqli jihatlarni ko‘rsatishga qaratiladi.

O‘zbek va turk parallel matnlar korpusidan o‘zbek tili tarixi, turkiy tillar qiyosiy-tarixiy grammatikasi, semantika fanlarini o‘qitishda asarlar mazmuni o‘zlashtirish, so‘zlar mazmunini yoritish, konseptual tahlil qilish, matn lingvistikasi asoslarini, har ikki tilga xos fonetik, grammatik, leksik xususiyatlarini o‘rganishda parallel matnlar korpusining ahamiyati katta.

O‘zbek-turk parallel matnlar korpusi umumturkiy yozma manbalarning qiyosiy planda o‘rganilishga zamin yaratadi. Ularning hozirgi o‘quvchilarga tushunarli bo‘lishini ta‘minlaydi. Bunda, albatta, tabdil va izohlardan foydalanish kerak bo‘ladi.

Badiiy asarlar asosidagi parallel matnlar korpusida frazemalar bilan bog‘liq muammolar bo‘lishi aniq. O‘zbek tilidagi frazemalar ikki yoki undan ortiq so‘zdan tuzilib, shu so‘zlar semantikasi asosida yangi ko‘chma lug‘aviy ma‘noni shakllantirishga xizmat qiladi. Shu bois parallel matnlar korpusini shakllantirishda to‘g‘ridan to‘g‘ri mashina tarjimasidan foydalanilsa, semantikasi to‘g‘ri yoritishda muammolar kelib chiqadi. Muammolarni bartaraf etish uchun frazemalar sodda va qo‘shma leksik birliklar, so‘z birikmalaridan farqlashi, maxsus teglar, razmetkalar bilan belgilanishi lozim. Frazemalar tarjimasida, albatta, mutaxassis nazorati kerak bo‘ladi.

Turk Milliy Korpusi zamonaviy turk tili uchun muvozanatli, keng ko‘lamli (50 million so‘z) umumiy maqsadli korpusdir. U korpusni qurishning oldingi amaliyotlari va harakatlaridan foyda ko‘rgan. Shu ma‘noda, TMK odatda Britaniya Milliy Korpusining tuzilishiga amal qiladi, ammo kerak bo‘lganda TMK tuzilishiga zarur tuzatishlar kiritiladi. Butun jarayon davomida aniq muammolarni hal qilish uchun ochiq kodli dasturiy ta‘minotning har xil turlari qo‘llaniladi va natijada olingan korpus notijorat maqsadlarda foydalanish uchun bepul resurs hisoblanadi. Turk tilining parallel korpusi yaratilgan.

## XULOSA

Parallel korpusni yaratish bir necha bosqichlarni o‘z ichiga oladi: matnni tekislash, matnni belgilash, qidiruv interfeysi dizayni.

Tartibga solish jarayoni, birinchi navbatda, tarjimada asl nusxaga mos keladigan fragment mavjudligini ta‘minlash uchun

ishlatiladi. Shundan soʻng, parallel matnlarning bir xil qismlari bir-biri bilan taqqoslanadi.

Dastlabki bosqichda paydo boʻladigan savol, aslida nimani tekislash kerak. Siz soʻzni soʻz bilan tekislashingiz mumkin, ammo bu koʻpincha bir qator sabablarga koʻra deyarli imkonsiz boʻlib chiqadi: tokenlar toʻplami, turli tillardagi barqaror iboralar mos kelmaydi. Shuningdek, tekstlar jumlar boʻyicha tekislanadi, ammo bu holda muammolar paydo boʻlishi mumkin: jumlar yoki paragraflar soni ham farq qilishi mumkin. Quyidagi oʻzbek tilidagi asl matnning turk tiliga qilingan qilingan tarjimasini koʻrishimiz mumkin. Soʻzlarni soʻzlar bilan tekislash muammo keltirishini oʻz isbotini topdi. Oʻzbek tilidagi soʻzlar soni turk tilidagi tarjima bilan uygʻun emasligini shu jumla orqali izohlaymiz. Oʻzbek tilida berilgan namunada 13 ta soʻz qatnashgan. Turk tiliga qilingan tarjimada esa 9 ta soʻz.

«Kambagʻal boʻlsang koʻchib boq, boyvachcha boʻlsang oʻlib boq», degan gapda hikmat koʻp ekan.

«Fakirsən taşınmayı dene, zenginsen ölmeyi dene» atasözünde hikmet varmış.

Bundan shu xulosaga kelinadiki jumla boʻyichamoslashtirish qulayroq hisoblanadi.

## REFERENCES

1. Abduraxmonova N.Z. "Oʻzbek tili elektron korpusining kompyuter modellari" filologiya fanlari doktori (DSc) dissertatsiyasi avtoreferati. Toshkent – 2021
2. Abduraxmonova, N. Z. "Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref." (2018).
3. Abdurakhmonova N. The bases of automatic morphological analysis for machine translation. Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta. 2016;2 (38):12-7.
4. Abdurakhmonova N, Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/Foreign Philology: Language. Literature, Education. 2018(3):68.
5. Abdurakhmonova N, Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL). 2019;6(1-2019):131-7.
6. <http://ruscorpora.ru>
7. <https://tanersezer.com/?p=308>
8. <https://dic.academic.ru/dic.nsf/ruwiki/1085886>

9. <https://translationrating.ru/glossary/parallelnyj-korpus/>
10. <https://huseyindemirtas.net/butunce-dilbilimi-corpus-linguistics-ve-ceviribilim/>
11. <https://ruscorpora.ru/new/search-para.html?lang=eng>
12. <http://www.nevmenandr.net/fr/>

