

Turk-o'zbek parallel korpusiga matn tanlashda muhim xususiyatlar

Xolmonova Iqbola Alisher qizi

ToshDO'TAU, kompyuter
lingvistikasi yo'nalishi 1-kurs magistranti
iqbolabintualisher@gmail.com

Ilmiy rahbar: Sh.Xamroyeva,

ToshDO'TAU dotsenti, filologiya fanlari
doktori

Annotatsiya: Ushbu maqolada Turk-o'zbek parallel korpusiga matn tanlashda muhim xususiyatlar haqida so'z boradi. Matn tanlashda kompilyatorning vazifalari sanab o'tiladi. Korpusga matn tanlashda qanday uslublardan foydalanish mumkinligi aytiladi. Bundan tashqari, maqolada korpusga matn tanlashda qadamlar ketma-ketligi sanab o'tilgan.

Kalit so'z: *matn, parallel korpus, avtomatik ekstraktsiya, gibrid yondashuv, matn janri, parallel matn.*

Annotation: This article discusses important features in text selection for the Turkish-Uzbek parallel corpus. The tasks of the compiler are listed in the text selection. It tells you what styles can be used when selecting text for the corpus. In addition, the article lists the sequence of steps in selecting text for the Corpus.

Keywords: *text, parallel corpus, automatic extraction, hybrid approach, text genre, parallel text.*

Аннотация: В данной статье рассматриваются важные особенности выделения текста для турецко-узбекского параллельного корпуса. Задачи компилятора перечислены в текстовом выделении. Он сообщает вам, какие стили можно использовать при выборе текста для корпуса. Кроме того, в статье указана последовательность действий при выборе текста для Корпуса.

Ключевые слова: *текст, параллельный корпус, автоматическое извлечение, гибридный подход, жанр текста, параллельный текст.*

Kirish. Korpus ma'lum prinsiplarga muvofiq yaratilgan katta matnlar to'plami hisoblanadi. Lekin bu har qanday matnlar to'plami degani emas, ya'ni korpusga kiritiladigan matnlar ma'lum talablarga javob berishi kerak.

Korpus matnlari kompilyatorning korpusni yaratish maqsadidan kelib chiqib, aniq mezonlar asosida tanlanadi. Kompilyator korpus uchun matnlarning hajmi, janri, turi va boshqa parametrlarini o'zi belgilaydi. Korpus tarkibidagi matnlarning mazmun xususiyati va turlari jihatidan bir-biridan farq qilishi davrning butun lug'at boyligini ochib berish nuqtayi nazaridan muhim ahamiyatga ega. Korpus tarkibiga kiruvchi matnlar turli janrlardan tashkil topgan bo'lsa, undan turli soha vakillari hamda keng omma foydalana oladi.

Ma'lumki, har qanday korpusning asosini uning birliklari tashkil qiladi. Odatda, korpusning reprezentativligini ta'minlash uchun korpus birligi sifatida turli janrdagi matnlar tanlanadi. Shuning uchun har qanday til korpusini yaratishda, birinchi navbatda, uning tarkibiy qismi bo'lgan matnlarni tanlash tamoyillarini aniqlash kerak.

Korpus tarkibiga kiritilayotgan matnlar kompilyator tomonidan erkin tanlanadi. Misol uchun, kompilyator badiiy asarlar korpusini tuzmoqchi bo'lsa, unda korpus tarkibiga hikoya, roman, qissa janridagi asarlarni kiritishi mumkin. Yoki ilmiy yo'nalishdagi korpus tuzmoqchi bo'lsa, korpus tarkibiga maqola, tezis, dissertatsiya, ilmiy asarlar kiritilishi mumkin. Nazmga qiziquvchi kompilyator (korpus tuzuvchi shaxs) faqat nazmiy asarlardan tarkib topgan korpus tuzishi ham mumkin.

Matn janri parallel korpus uchun muhim xususiyatdir, chunki u matnlarda ishlatiladigan tilga ta'sir qiladi. Turli janrlarda turlicha konvensiya va uslublar mavjud bo'lib, ular tilning so'z boyligi, grammatikasi va tuzilishiga o'z ta'sirini o'tkazadi. Misol uchun, tasviriy til va adabiy vositalardan foydalanish mumkin bo'lgan romandan ko'ra ko'proq rasmiy til va texnik atamalardan foydalanish parallel matn uchun qulayroq. Parallel korpusga turli janrlarni kiritish orqali tadqiqotchilar tilning turli kontekst va janrlarda qanday farqlanishini va matn janriga qarab tarjimalar qanday farqlanishini tahlil qilishlari mumkin. Bu turli soha va kontekstlarda tildan foydalanish va tarjima amaliyotlari haqida qimmatli tushunchalar berishga xizmat qiladi [1. 111-122].

Asosiy qism. Matn tanlashda korpusning turi ham muhim ahamiyatga ega. Tadqiqotimizning maqsadi turk va o'zbek badiiy asarlarining parallel korpusi uchun matn tanlash muammosi bo'lganligi sababli ushbu maqolada asosan bu haqida so'z boradi. Aslida, parallel korpus nima uchun kerak? Bu savolga javoblar quyidagicha:

- 1) mavjud dunyo parallel korpusi tayyor tarjimalarga asoslanadi;
- 2) tarjimani parallel dasturga birlashtirish asl asarni va butun asar tarjimasini yonma-yon o'qish imkonini beradi;
- 3) eng muhimi, parallel korpus u yoki bu birlik (sodda gap, qo'shma gap, ibora, frazeologik birlik, jargon, metafora) asl nusxada va tarjimada qanday ifodalanganligini kuzatishning eng maqbul usuli hisoblanadi [2. 348].

Parallel korpus uchun matn tanlayotganda asl matn va uning tarjimasi kerak bo'lishini hisobga olish zarur. Shuning uchun kompilyatorning vazifasi matn tanlash jarayonida korpusga kiritmoqchi bo'lgan matning tarjimasi mavjud ekanligiga amin bo'lgandagina uni korpusga kiritishi mumkin bo'ladi.

Parallel korpus uchun matn tanlashda quyidagi usullardan foydalanish mumkin:

1. **Avtomatik ekstraktsiya.** Bu veb-sahifalar kabi yirik onlayn resurslardan parallel matnlarni olish uchun dasturiy vositalardan foydalanishni o'z ichiga oladi.
2. **Qo'lda tanlash.** Bu mutaxassisdan uzunlik, mazmun va janr kabi talablarga mos keladigan matnlarni qo'lda tanlashni talab qiladi.

3. **Gibrid yondashuv.** Bu zarur mezonlarga javob beradigan parallel korpusni yaratish uchun avtomatlashtirilgan vositalardan foydalanish va qo'lda tanlashni birlashtiradi [3. 154].

Parallel korpus uchun matn tanlash jarayonida bir qancha amallarni bajarish kerak bo'ladi:

1. **Tillarni aniqlash.** Parallel korpus uchun manba va maqsadli tillarni belgilash.

2. **Maqsadni aniqlash.** Parallel korpusdan maqsadli foydalanishni aniqlash kerak bo'ladi, jumladan, mashina tarjimasi, tilni modellashtirish va ikki tilli lug'at yaratish yoki chet tilini parallel ravishda o'rganish kabi.

3. **Mavjudligini ko'rib chiqish.** Tanlangan tillar uchun matnlar mavjudligini baholash va mavjud matnlardan foydalanish yoki yangilarini yaratish haqida qaror qabul qilish lozim.

4. **Janrlarni tanlash.** Badiiy matn, ilmiy matn, rasmiy matn va ijtimoiy media postlari kabi parallel korpus maqsadiga eng mos keladigan janrlarni aniqlash va tanlash amalini bajarish zarur.

5. **Matnlarni tanlash.** Uzunlik, izchillik va dolzarblik talablariga javob beradigan matnlarni tanlash uchun avtomatlashtirilgan dasturiy vositalar yoki ekspert inson baholovchilaridan foydalanish kerak. Matnlar tarjima sifatida bir-biriga mos kelishi, o'xshash mazmun va tuzilishga ega bo'lishi hamda muhim madaniy yoki lingvistik farqlarga yo'l qo'ymasligi kerak.

6. **Sifat nazoratidan o'tkazish.** Olingan parallel korpusni mo'ljallangan maqsad va sifat standartlariga javob berishini tekshirish lozim.[4. 238-264]

Parallel korpus birliklari asl nusxadan bir yoki bir necha marta tarjima qilingan matnlardan iborat bo'ladi. Parallel korpuslar bir tomonlama tarjimani (faqat o'zbek tilidan turk tiliga) yoki ikki tomonlama tarjimani (o'zbek tilidan turk tiliga, turk tilidan o'zbek tiliga) o'z ichiga oladi. Shundan kelib chiqib, o'zbek-turk parallel korpusi ikkita blogdan iborat bo'ladi:

1. O'zbek-turk tilidagi blog;

2. Turk-o'zbek tilidagi blog.

Bu bloglar uchun materiallar bir xil emas: birinchi blogda o'zbek tilidan turk tiliga tarjima qilingan badiiy asarlar mavjud bo'lsa, ikkinchi blog turk tilidan o'zbek tiliga tarjima qilingan asarlardan iborat. Albatta, qidiruv tizimida janr, yaratilish davri, asar muallifi va tarjimoni bo'yicha qidiruvni amalga oshirish mumkin bo'ladi.

Turk va o'zbek tilidagi matnlarning parallel korpusga kiritilishi bir qancha sabablarga ko'ra muhim ahamiyatga ega. Birinchidan, turk va o'zbek tillari o'z mintaqalarida keng tarqalgan tildir: turk tili Turkiyaning rasmiy tili, o'zbek tili esa O'zbekistonning davlat tilidir. Bundan kelib chiqadiki, ushbu tillarda juda ko'p yozma materiallar mavjud, ular parallel korpus qurish uchun qimmatli manbalarga aylanadi. Ikkinchidan, turk va o'zbek tilidagi matnlarning kiritilishi tadqiqotchilarga tilning turli til oilalari bo'yicha o'zgarishini tahlil qilish imkonini beradi. Qarindosh tillar o'rtasidagi tarjimalarni solishtirish orqali tadqiqotchilar lingvistik tuzilmalar va

xususiyatlarning turli til guruhlariga bo'ylab qanday taqsimlanishi yoki farqlanishi haqida tushunchaga ega bo'lishlari mumkin.

Nihoyat, turk va o'zbek tilidagi matnlarning parallel korpusga kiritilishi ushbu tillardagi tarjima amaliyoti haqida qimmatli fikrlarni beradi. Turk va o'zbek tilining o'ziga xos grammatik tuzilishi, so'z boyligi va idiomatik iboralari mavjud bo'lib, bu tarjimonlar uchun qiyinchilik tug'dirishi mumkin. Ushbu tillar va korpusdagi boshqa tillar o'rtasidagi tarjimalarni tahlil qilish orqali tadqiqotchi, tarjimonlar bu qiyinchiliklarni qanday hal qilishlari va o'z tarjimalarini turli janr, kontekstlarga qanday moslashtirishlari haqida tushunchaga ega bo'lishlari mumkin bo'ladi.

Xulosa qilib aytganda, turk va o'zbek tilidagi matnlarning parallel korpusga kiritilishi korpusni yaratish uchun qimmatli manbalarni taqdim etishi, qarindosh tillararo tahlil qilish imkonini beradi, mazkur tillardagi tarjima amaliyoti haqida tushuncha berishi bilan ahamiyatli.

Foydalanilgan adabiyotlar:

1. Trosterud, Trond 2002: Parallel corpora as tools for investigating and developing minority languages. In: Borin, Lars. Parallel Corpora, Parallel Worlds. Selected papers from a symposium on parallel and comparable corpora at Uppsala University, Sweden, 22-23 April 1999. Amsterdam New York, NY: Rodopi. Pp. 111-122.
2. Михайлов М. Параллельные корпуса художественных текстов: диссертация. Финляндия. ун-та, 2003. 348с
3. McEnery, Tony and Wilson, Andrew 2001: Corpus Linguistics: An Introduction. 2nd edition. Edinburgh: Edinburgh University Press.
4. Salkie, Raphael 2002: how can linguists profit from parallel corpora? In: Borin, Lars (ed.). Parallel Corpora, Parallel Worlds. Selected papers from a symposium on parallel and comparable corpora at Uppsala University, Sweden, 22-23 April 1999. Amsterdam New York, NY: Rodopi. Pp. 93-10
5. Teubert, Wolfgang 1996: Comparable or Parallel Corpora. International Journal of Lexicography. Oxford University Press. 9 (3), 238-264.
6. Elov B., Amirqulov M. O'zbek-ingliz tillarining teglangan parallel korpusini yaratish bosqichlari. O'zbekiston: til va madaniyat (Amaliy filologiya), 2022, 5(4). 90-102.
7. Hamroyeva Sh. O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Filol. Fan. bo'yicha falsafa dokt. (PhD) diss. - Buxoro, 2018. B-73.
8. Abdullayeva O. O'zbek tili korpusini yaratish bosqichlari va muammoli jihatlari. / "So'z san'ati", 4-jild, 1-son – Toshkent, 2021. –b. 153-158.
9. Abdullayeva O., Norqulova M. Internet matnlarining til korpusi rivojida ahamiyati. Science and innovation. Vol. 1, Issue 6. – Toshkent, 2022. (doi.org/10.5281/zenodo.718910).