

O'ZBEK TILI PARALLEL KORPUSINING LINGVISTIK MASALALARI

Abduraxmonova Nilufar Zaynobiddin qizi

filologiya fanlari doktori, professor
O'zMU Kompyuter lingvistikasi va amaliy
tilshunoslik kafedrası mudiri

Vasliddinova Kamola

O'zDJTU tayanch doktoranti

Annotatsiya. Tabiiy tilni qayta ishlash jarayoni uchun korpus texnologiyasini rivojlantirish dolzarb masalalardan biridir. Mashina tarjimasi, savol javob tizimi, tarjima nazariyasi va amaliyoti uchun parallel korpus texnologiyasini yaratishda lingvistik baza muhim resurs hisoblanadi. O'zbek tili korpusi uchun ko'p tilli parallel korpus yaratishning lingvistik masalalarni tizimli tadqiq etish va til bilan bog'liq leksikografik tadqiqotlar haqidagi nazariy mulohazalar ushbu maqolada tahlilga tortilgan.

Abstract. Development of corpus technology for natural language processing is one of the important issues. A linguistic database is a significant resource in building parallel corpus technology for machine translation, response systems, and theory and applied translation. Systematic research of linguistic issues of creating a multilingual parallel corpus for the Uzbek language corpus and theoretical considerations of language-related lexicographic research are analyzed in this article.

Аннотация. Развитие корпусной технологии обработки естественного языка является одним из важных вопросов. Лингвистическая база данных является важным ресурсом в создании технологии параллельных корпусов для машинного перевода, систем реагирования, а также теории и практики перевода. Лингвистическая база данных является важным ресурсом в создании технологии параллельных корпусов для машинного перевода, систем реагирования, а также теоретического и прикладного перевода. В статье анализируются системное исследование лингвистических вопросов создания многоязычного параллельного корпуса корпуса узбекского языка и теоретические соображения языковых лексикографических исследований.

Kalit so'zlar: *parallel korpus, kompyuter leksikografiyasi, mashina tarjimasi tizimlari, korpusga asoslangan terminologik ma'lumotlar bazasi.*

Bizning tadqiqotimiz uchun asos bo'luvchi korpus strukturasi va uning kontentida parallel korpuslar ham o'rin olgan. Shu bois parallel korpus va uni o'zbek tili uchun yaratish masalalari ishimiz asosini tashkil qiladi. Ma'lumki, korpus faqat bir til doirasida yaratilgan bo'lsa, monolingval korpus, boshqa tillar doirasida yaratilgan bo'lsa parallel korpus deb nomlanadi. Parallel korpusda asliyatdagi matn

va uning muayyan tildagi tarjimasiga doir *paragraf, jumla yoki so'zlarning o'zaro muqobil juftliklari* bazaga kiritiladi.

Parallel korpus termini ostida mualliflar tomonidan muayyan tavsifiy xarakterdagi turlarga egaligi qayd etiladi [Kudela, 2016]:

- parallel korpus bir nechta tillardagi hujjatlarning sifatli tarjimalarini o'z ichiga olgan korpus hisoblanadi. Bunday korpuslarda tarjima muqobillari aniqlanib, gap darajasida ularni moslash nisbatan oson;

- “shovqinli” parallel korpus – bu u darajada sifatli tarjimaga ega bo'lmagan hamda tarjima muqobillari ham mukammal darajada muvofiqlashmagan bo'ladi. Shunday bo'lsa-da, ulardagi tarjimalarni korpus tarkibiga kiritish mumkin;

- qiyosiy parallel korpus – mavzu jihatdan mushtarak, biroq aynan matnning sinatktik va semantik jihatdan bir-biridan farqli bo'lgan korpus turiga tegishlidir. Ushbu holatda hujjatlar mavzu jihatdan umumiy guruhga tegishli bo'ladi. Muayyan mazmundagi asliyatdagi matn va uning tarjimalari baza sifatida korpusga kiritiladi. Masalan, qiyosiy korpus sifatida olingan turli OAVdagi yangiliklar muayyan vaqtga tegishli voqealarni aks ettiradi. Ushbu holatda matnlarning o'zaro muqobillik belgilari emas, balki so'zlarni mavzuni yoritishda qo'llash mahorati va matn strukturasini shakllantirish uchun qiyosiy-chog'ishtiruv aspektida o'rganiladigan sohalarning obyekti vazifasini bajaradi;

- kvazi-qiyoslash mumkin bo'lgan korpus – bu xilma-xil mavzudagi va grammatik jihatdan o'zaro parallel bo'lmagan hujjatlarni o'z ichiga oladi.

Opus platformasida ingliz-shved qiyosiy korpusi wikipediya olingan ma'lumotlar bo'yicha yaratilgan. Shuningdek, dunyoning turli mintaqalarida ingliz tilida so'zlashuvchi Britaniya ingliz tili, Avstraliya yoki hind ingliz tilining bir million so'zlik subkorpusdan iborat xalqaro korpusi bir xil namuna asosida jamlangani bois u *qiyosiy korpus* sifatida qabul qilinadi [Kubler, Zinsmeister, 2015].

Parallel korpusning mazmuniy tomonini hisobga olib, V.Zaxarov mazkur korpuslarni shartli ravishda ikkiga ajratadi:

- i.muayyan tildagi asliyat matnning ikkinchi tildagi aynan tarjimas;

- ii.biror sohaga oid mavzulashtirilgan guruhlar bo'yicha bir tasnifga kiritilgan matnlar.

Bugungi kunda parallel korpus yaratishning avtomatik texnologiyalari ishlab chiqilgan bo'lib, ularning aksariyati Yevropa tillari uchun instrumentariy vazifasini bajaradi. Shular qatoriga web sahifalardagi korpuslari generatsiya qiluvchi *Bitextor* instrumentariysini keltirish mumkin [Abdurakhmonova, Alisher, Toirova, 2022]. U yaxlit ravishda web-saytni yuklab olib, fayllardagi tillarni aniqlaydi va o'xshash mazmundagi matnlarni muvofiqlashtiradi, so'ng ularni html yoki xml formatda bazaga saqlaydi. Hujjatlar hajmi va muqobil tarjima birliklari bo'yicha tarjima xotirasiga (Translation Memory eXchange -TMX) uzatiladi [Kudela, 2016]. Bitextorning keyingi versiyalari yanada takomillashtirilib, URL (Uniform Resource Locator) va so'zlar to'plami (bag-of-words) modeli asosida filtrlash imkoniyatiga ham ega bo'ldi [Abdurakhmonova, Tuliye, Ismailov, Abdurahobov, 2022].

Avtomatik usulda parallel korpus yaratishning yana bir turi *PaCo2* instrumentariysi orqali ko'rsatish mumkin. Mazkur texnologiya internetdan parallel korpuslarni yig'ishga mo'ljallangan to'liq avtomatlashtirilgan tizim hisoblanadi [Kudela, 2016]. U ham tanlagan tillar bo'yicha bilingval kontentli web-saytlardan ma'lumotlarni yuklash imkoniyatiga ega. Tilni tanlash ixtiyoriy bo'lgani uchun ham ushbu instrument barcha tillar uchun moslasha oladi [Abdurakhmonova, 2019]. U uch asosiy bosqichga ega: 1- bosqichda bilingval kontentli web-saytlarni muayyan til bo'yicha qidiradi. 2- bosqichda dastlabki bosqichda aniqlangan saytdagi parallel web-sahifalar ajratiladi. So'nggi bosqichda parallel web-sahifalar parallel korpus sifatida har bir gapning muqobil juftlari bo'yicha guruhlanadi.

STRAND parallel korpus instrumentariysi hisoblanib, tarjimalarni strukturaviy aniqlash imkoniyatiga ega tizim hisoblanadi [Kudela, 2016]. Ushbu texnologiya web-sahifalardagi parallel juftlarini aniqlashga qaratilgan. U oddiy saytlardagi mazmuniy o'xshashligi bo'lgan matnlarni avtomatik tahlil qiladi

Bugungi kunda mashina tarjimasi tizimlari uchun ochiq platformali parallel korpus bazalari yaratilmoqda. Opus parallel korpus platformasi (<http://opus.nlpl.eu/>) lingvistik resurslar, korpus interfeysi va parallel matnlar bazasidan tashkil topgan. Qayd etish lozimki, ilk bor o'zbek tili uchun *Tanzil* deb nomlangan elektron resursda Qur'oni Karimning o'zbek tilidagi tavsiri parallel matn sifatida bazaga kiritilgan [Abdurakhmonova, Tuliyeu, Gatiatullin, 2019].

Shu o'rinda *CLARIN* *infrastrukturasi*da 86 turdagi bilingval korpuslarning bazasi yaratilganini alohida qayd etish o'rinli. Undagi mavjud parallel korpuslar Yevropa va boshqa til oilalariga mansub til juftliklarining *hajmi*, *lingvistik razmetkasi*, *foydalanish imkoniyati*, *muqobillik konkordansi*, *matnlarning uslubiy ko'lami va undagi janrlar xilma-xilligiga* ko'ra farqlanadi.

Ochiq platformali parallel matnlar muhitini yaratish uchun CAT (computer assisted translation- kompyuter yordamidagi tarjima) texnologiyasi yordamida matnlarga ishlov berishning dastlabki bosqichi amalga oshiriladi. Mashina tarjimasining CAT texnologiyasi parallel matnlar asosida lingvistik resurs yaratishning eng maqbul usuli sifatida dunyo olimlari tomonidan tan olingan. Tarjima xotirasi muhiti yordamida ikki tildagi gaplarning o'zaro *muqobillik (alignment)* holati aniqlanadi, tarjima instrumenti SDL sifatida foydalaniladi. Instrumentariylar yordamida tarjimon "tarjima xotirasi" bankini yaratadi. Natijada professional tarjimon muayyan sohada tarjimani amalga oshirishda tarjima qilingan matnlarni qayta ishlatishi, avval tarjima qilingan segment birliklarni mos tartibga ajratish bosqichi avtomatik tarzda amalga oshirishi, o'xshash tarjimalar yoki terminlarni ilova qilish imkoniyatiga ega bo'ladi. Tarjima instrumentariylari onlayn va kompyuter platformasiga moslashishi lozim. Kompyuter platformasi uchun pulli foydalanishga mo'ljallangan *Trados studio*, *memoQ*, *Wordfast*, *Memsources*, *Déjà Vu*, *Across*, bepul xizmatlardan *Omega T*, *Café TranEspresso*, pulli onlayn xizmatlardan *Memsources*, *Wordbee*, *XTM Cloud*, tekin foydalanish imkoniyatiga ega *Smartcat*, *Matecat* va *Wordfast Anywhere* kabilarni keltirish mumkin.

Tarjimonlik sohasida ham bir qancha lingvistik instrumentlar mavjud. Ular parallel tarzda tarjima muqobillarini serverda to'plash hamda bilingval terminologik ma'lumotlar bazasini yig'ishga mo'ljallangan. Parallel korpusga asoslangan bilingval terminlarni boyitib borish o'tgan asrning 80-yillariga borib taqaladi. Shu davrda yaponiyalik olim Makato Nagao namunaga asoslangan (example based translation) mashina tarjimasi texnologiyasini ilgari surgan. Endilikda ushbu turdagi texnologiyalar tarjima xotirasi (*TM-translation memory*) nomini oldi.

Korpusga asoslangan terminologik ma'lumotlar bazasini yaratish internet tizimidagi ma'lumotlarning parallel ravishda berilishi hamda mehnat resursining tejamkorligi jihatdan yuqori bo'lganligi bois g'arbda tadqiqot usuli sifatida an'anaga aylangan mazkur texnologiyadan maqsadli foydalanib kelinmoqda. Ushbu texnologiya mashina tarjimasi uchun mehnat samaradorligi va uni izchil ravishda boyitib borish imkoniyatiga egaligi bilan ahamiyatlidir. Shunday instrumentlardan biri *Paraconc* texnologiyasidir. U bilingval va ko'p tilli dastur bo'lib, quyidagi imkoniyatlarga ega [Ketiwe, <http://www.literator.org.za>]:

1. Turli usullarda so'zlarni saralash va hisoblash;
2. Tarjima qilingan parallel matnlarni moslashtirish;
3. Matndagi mavjud birliklarni oson topish va ularni ko'rsatish;
 - 1) terminlarni qidirish;
 - 2) tarjima ekvivalentlarini aniqlash;
 - 3) tarjimalardagi muhim o'rinlarni aniqlash;
 - 4) so'z birikmasi sifatida so'zlar o'ramini tahlil qilish;
 - 5) so'zlar chastotasi.

Kompyuter yordamidagi mashina tarjimasi texnologiyasi (CAT –computer assisted translation) ikki qismga ajralgan:

- 1) *Tarjima xotirasi* (Translation memory)
- 2) *Terminologik boshqaruv* (Terminology management)

Parallel matnlar asosida korpus yaratish u darajada oson emas. Chunki ikki tilning mos matn fragmentlari o'zaro juftlanadi. Ushbu masalani avtomatik tarzda hal qilish uchun turli metodlar ishlab chiqilgan, buning uchun ikki tildagi matndagi grammatik konstruktsiya (gaplar va so'z birikmalari), so'zlar muvofiqlashtiriladi. A.V.Zubov va I.I.Zubovalarning darsligida gaplar darajasida gaplarning muqobilligini aniqlash bo'yicha fikrlar o'rtaga tashlangan. Ushbu moslashtirish olti usulda amalga oshirilishi mumkin [3axapob, 2013]:

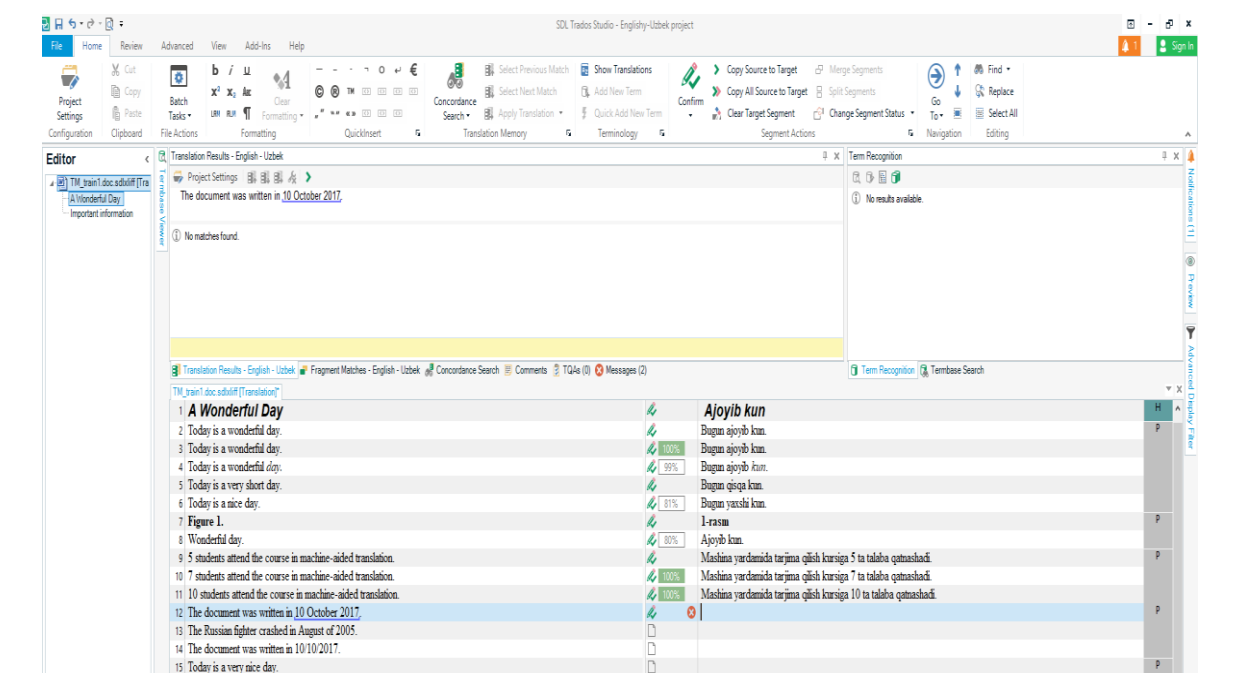
- 1) asliyatdagi bir matn ikkinchi tilga aynan hajmda tarjima qilinishi;
- 2) asliyatdagi ikki matn ikkinchi tilga bir matn bo'lib tarjima qilinishi;
- 3) asliyatdagi bir matn ikkinchi tilga ikkita matn sifatida tarjima qilinishi;
- 4) asliyatdagi ikki gap ikkinchi tilga ham ikkita jumla tarzida tarjima qilinishi;
- 5) asliyatdagi jumla ikkinchi tilga umuman tarjima qilinmasligi;
- 6) asliyatdagi matnning tarjima qilinadigan tilda mos ekvivalentining mavjud emasligi.

Parallel matnlarning mos juftliklarini tuzish uchun inson tomonidan tahrir qilingan matnlar kiritilib, lug‘at orqali matnda uchragan segment birliklarga qarab avtomatik tarzda tarjima birliklari aniqlanadi. Buning uchun ikki tilga mos leksik va grammatik jihatdan ishlab chiqilgan variantlar lug‘atga kiritilgan bo‘lishi kerak. Parallel korpus quyidagi maqsadlarda qo‘llanilishi mumkin [Захаров, 2013]:

- ikkit yoki ko‘p tilli tarjima lug‘atlar tuzishda;
- mashina tarjima tizimlari uchun lug‘at yaratish va uni doimiy tarzda to‘ldirib borish;
- kontekstda uchraydigan ko‘p ma’noli so‘zlarning kompyuter tahlili orqali leksik birliklarni polisemiya o‘ldirishini aniqlash;
- matndagi terminologik va frazeologik birliklarni tarjima qilish;
- korpusdan foydalangan holda tarjimalarning mos variantlarini kompyuter xotirasiga yuklash va shu orqali mashina tarjimasi tizimi uchun to‘liq avtomatik tarjima amalga oshirish.

Leksikani avtomatik tasniflash matnni avtomatik tushunish uchun asosiy amallardan biri hisoblanadi [Беляева, 2004].

Tarjima instrumentlariga tarjima xotirasi texnologiyasiga (translation memory) asoslangan Trados dasturi hamda WinAlign tarjima texnologiyasi kiradi (CAT –computer assisted translation). Ushbu dasturlar *Translator’s workbench* va *MultiTerm* bazasi asosida yaratilgan. Bunda jumlar sifatida nuqta yoki gap uchun ishlatiladigan boshqa ishoraviy belgilargacha bo‘lgan qator gap sifatida segmentlarga ajratiladi [Abdurakhmonova va b., 2022]. Professional tarjimonlar orqali qilingan tarjima matnlar parallel ravishda kompyuter xotirasiga yuklanadi. Keyingi safar avvalgi qilingan tarjimaga o‘xshash matn yoki ibora uchrasa, kompyuter foiz ko‘rsatkichida segment o‘xshashlikni ilova qiladi. Bu esa tarjimonga muayyan yo‘nalishda tarjima qilinayotgan asarni qayta tarjima qilmay, bazada



mavjud ekvivalentlikni topish ixtiyorini voqelantiradi. Bunda ma'lumotlarni foydalanuvchilar tomonidan boyitib borilgan tarjimalar hisobiga kompyuter xotirasidagi asos va tarjima til uchun yaratilgan tayyor tarjima qoliplari hosil bo'ladi. Manbalarda qayd etishicha, ushbu tarjima uskunlaridan foydalanib inglizcha ilmiy matnlar nemis tiliga tarjima qilinganda, 28% tezroq mehnat samaradorligiga erishish mumkin [Marcos, Mihaela, 2014].

Quyida Trados dasturi orqali inglizcha matnni o'zbek tiliga tarjima qilish namunasi ko'rsatilgan (1-rasm). Tarjima xotiraning afzallik jihat – bir vaqtning o'zida tarjima qilingan so'zlar bilingval lug'at sifatida xotirada saqlanadi. So'ngra undan kontekstga asoslangan tarjima lug'at sifatida foydalanish mumkin. Tradosning yana bir ahamiyatli jhati – tarjima qilinayotgan matn .doc hamda HTML, SGML, XML, ASP, JSP, Tagged RTF, TRADOSTag kabi formatlarni ham o'qishga mo'ljallangan *TagEditor* dasturining mavjudligidadir. Bu esa asarni ayni aslyatdagi shaklida tarjima qilish imkoniyatini yaratadi

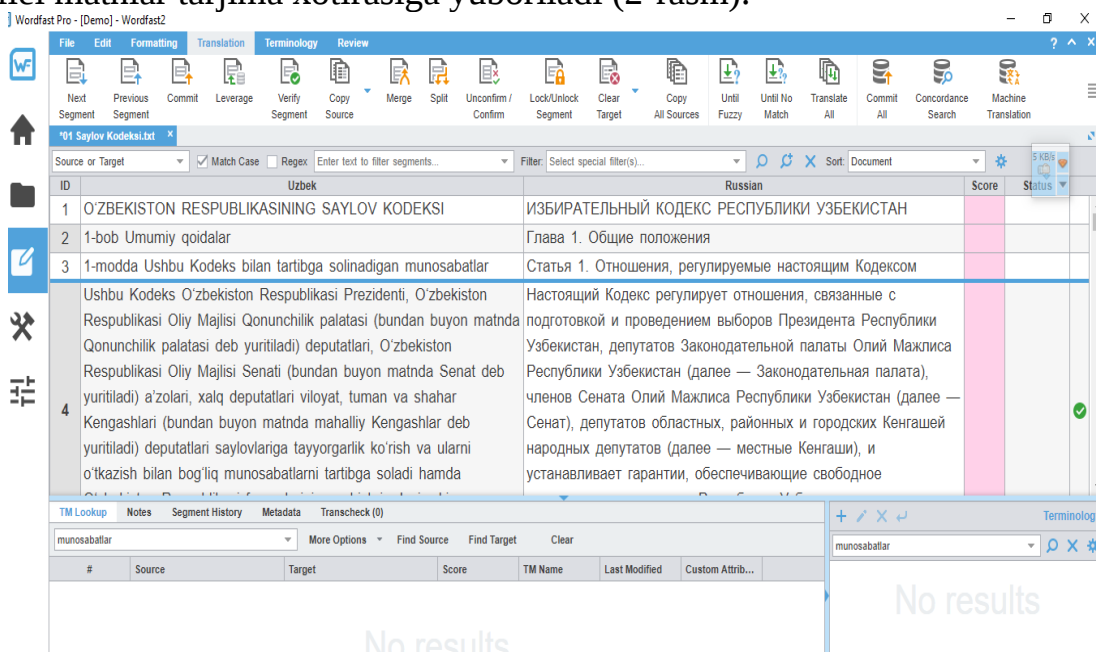
Mazkur tadqiqotda Wordfast Pro versiyasidan foydalanib, malakali tarjimonlar tomonidan o'zbek tilidan rus va ingliz tiliga tarjima qilingan matnlarning elektron shakldagi (.txt formatda) yuzdan ortiq ilmiy va rasmiy uslubdagi [<https://president.uz/uz> va <http://ziyonet.uz/>] parallel matnlar jamlandi (1-rasm).

ID	Russian ()	Score
1	ИЗБИРАТЕЛЬНЫЙ КОДЕКС РЕСПУБЛИКИ УЗБЕКИСТАН	-1
2	Глава 1. Общие положения	-1
3	Статья 1. Отношения, регулируемые настоящим Кодексом	-1
4	Настоящий Кодекс регулирует отношения, связанные с подготовкой и проведением выборов Президента Республики Узбекистан, депутатов Законодательной палаты Олий Мажлиса Республики Узбекистан (далее — Законодательная палата), членов Сената Олий Мажлиса Республики Узбекистан (далее — Сенат), депутатов областных, районных и городских Кенгашей народных депутатов (далее — местные Кенгаши), и устанавливает гарантии, обеспечивающие свободное волеизъявление граждан Республики Узбекистан.	-1
5	Статья 2. Законодательство о выборах	-1
6	Законодательство о выборах состоит из Конституции Республики Узбекистан, настоящего Кодекса и иных актов законодательства Республики Узбекистан.	-1

Ushbu matnlar tibbiyot, matematika, axborot xavfsizligi, tarjima, siyosat, filologiya kabi turli sohalarini o'z ichiga qamrab olgan. Ma'lumot o'rnida shuni aytish mumkinki, internetning imkoniyatlaridan unumli foydalanib, Webcrawler yordamida ko'p tilli saytlar yoki wikipediylarni ham tarjima xotirasiga yuklash mumkin. Manbalarda ta'kidlanishicha, yuklab olingan bilingval saytdagi muqobil parallel matnlar avtomatik generator Betextor yordamida TMX faylga to'planadi [http://lrec-conf.org/workshops/lrec2018/W3/pdf/5_W3.pdf]. Biroq generator faqat kam ta'minlangan resursli tillar uchun amal qilib, turkiy tillarning boshqa til juftliklari uchun mo'ljallangan emas.

O'zbek tilining o'ziga xos xususiyatlarini hisobga olib, internetda mavjud portallardan sohaga doir parallel matnlarni to'plash jarayonning dastlabki bosqichlaridan biri sanaladi.

Tarjimaning nechog'lik sifatli tarjima qilinishi so'z, so'z birikmasi yoki iboralarning uslub jihatdan to'g'ri tanlanganligi tahlil qilinib, segmentlash yoki tarjima jarayonida yo'l qo'yilgan xatolar tarjimon va tilshunoslar tomonidan qayta tahrirlanish jarayoniga ham bog'liqdir. So'z konkordanslarini aniqlash va ilmiy uslubdagi matnda qo'llangan terminlar uchun maxsus dastur yordamida ma'lumotlar bazasining shakllantirilishi, ajratib olingan parallel matnlardagi til va nutqiy birliklar alohida jadval shaklga keltirilishi tarjima dasturi uchun dastlabki bosqichlardan biridir. Segmentlash jarayoni to'liq yakuniga yetkanidan so'ng barcha parallel matnlar tarjima xotirasiga yuboriladi (2-rasm).



Shuningdek, ajratilgan gap segmentlaridan terminlar bazasi, so'z va so'z birikmalari avtomatik ravishda aniqlanib, inson tomonidan uning muqobillik darajasi tahrir qilinadi.

Ilmiy matnlardagi doimiy ishlatiladigan me'yoriy qoliplar sifatida quyidagi kalit so'zlarga e'tibor beriladi: so'z *asosi* => *derivatsion shakli* => *so'z shakli* => *ko'p qo'llanilgan so'z, fraza yoki terminlarning chastotasi* => *konkordansi* => *tarjimasi*. Namunada ko'rsatilgan matn “uzun” jumlati matn turiga kiradi, chunki unda tinish belgilar, qolaversa, predikativ elementlar miqdori ko'p bo'lib, ushbu matn tarjimon tomonidan ikki qismga ajratilib tarjima qilingan. Yirik matnli matndan ajratib olingan parallel jumalar generatsiyasi quyidagi bosqichlarda amalga oshirilishi mumkin [Копотев, 2014]:

I. Parallel matndan olingan tokenlashgan so'zning muqobil juftlari haqida ma'lumot olinadi;

II. Uzun jumlati matn muayyan tinish belgilari “,”, “;”, “:” bo'lgan qismigacha alohida segment gaplarga ajratiladi;

III. Asos – tarjima qilingan gaplarning segmentlari tarkibining miqdori aniqlanib, muvofiqlik holati belgilanadi;

IV. Tarjima – asos matnlarning muqobillik holati aniqlanadi;

V. Segmentlar o'rtasida birga ko'p yoki ko'pga ko'p munosabati kuzatilganda birga bir munosabat shaklida bir nechta segment birliklar o'zaro bog'lanadi yoki biriktiriladi.

Parallel korpuslar yordamida so'z va so'z birikmalarining o'zaro valentliklarini belgilash muayyan guruhga kiruvchi leksik-semantik guruhdagi so'zlarning semantik maydonini aniqlash, qolaversa, mashina tarjimasini uchun tarjima sifatini oshirishga hizmat qiladi.

Foydalanilgan adabiyotlar:

1. Kudela J. Mining Parallel Corpora from the Web. (master's thesis). Prague, 2016. – P. 13.
2. Kübler S., Zinsmeister H. Corpus linguistics and linguistically annotated corpora. – New York: Bloomsbury Academic, 2015. - P. 21.
3. Kudela J. Mining Parallel Corpora from the Web. (master's thesis). Prague, 2016. – P.4 .
4. Kudela J. Mining Parallel Corpora from the Web. (master's thesis). Prague, 2016. – P. 7.
5. Ketive N. Using ParaConc to extract bilingual terminology from parallel corpora: A case of English and Ndebele Literature / Journal of Literary Criticism, Comparative Linguistics and Literary Studies ISSN: (Online) 2219-8237, 0258-2279 <http://www.literator.org.za>
6. Захаров В.П., Богданова С.Ю. КОРПУСНАЯ ЛИНГВИСТИКА. (Учебник) - Санкт-Петербург, 2013, - С. 20.
7. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.
8. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In *компьютерная обработка тюркских языков. Turklang 2022* (pp. 231-240).
9. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In *of the International Conference on Language Technologies for All (LT4All)*.
10. Abdurakhmonova, N., Tuliyeu, U., & Gatiatullin, A. (2021, November). Linguistic functionality of Uzbek Electron Corpus: [uzbekcorpus. uz](http://uzbekcorpus.uz). In *2021 International Conference on Information Science and Communications Technologies (ICISCT)* (pp. 1-4). IEEE.
11. Abdurakhmonova, N., Alisher, I., & Toirova, G. (2022, September). Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. In *2022 7th International Conference on Computer Science and Engineering (UBMK)* (pp. 73-75). IEEE.



12. Беляева Л.Н. Лексикографический потенциал параллельного корпуса текстов // Труды международной конференции «Корпусная лингвистика – 2004» – СПб., 2004. – С. 55-64.
13. Marcos Z., Mihaela V. Quantifying the Influence of MT Output in the Translators' Performance: A Case Study in Technical Translation / Workshop on Humans and Computer-assisted Translation, Gothenburg, Sweden, 2014. – P. 93–98.
14. Копотев М. Введение в корпусную лингвистику. – Прага, 2014. – С. 129-132.
15. <https://president.uz/uz> va <http://ziyonet.uz/> manzilidagi yozma materiallar