

UDK: 811.512.133`42

LINGVISTIK TADQIQOTLARDA YANGI YONDASHUV: ABDULLA QODIRIYNING “O‘TKAN KUNLAR” ROMANI ASOSIDA YARATILGAN PARALLEL KORPUSI VA UNING TAVSIFI

O‘tkirova Fotima Baxtiyor qizi,
Kompyuter lingvistikasi mutaxassisligi magistranti
otkirovafotima@gmail.com
ToshDO‘TAU

Annotatsiya. Ushbu maqolada korpus lingvistikasi sohasining bir qismi bo‘lgan parallel korpus va uning tilshunoslikdagi ahamiyati haqida so‘z boradi. Lingvistika va adabiyot o‘rtasidagi ilmiy izlanish va muammolarga yangicha yondashuv orqali yechim berish to‘g‘risida ma’lumot beriladi. Abdulla Qodiriyning “O‘tkan kunlar” romani asosida yaratilgan o‘zbek-rus hamda rus-o‘zbek parallel korpusi tavsifi yoritiladi. Asarning har ikki korpusda ifodalangan til imkoniyatlari, muallif hamda tarjimon qalamiga xos xususiyatlar ko‘rsatiladi.

Abstract. This article discusses the parallel corpus, which is part of the field of corpus linguistics, and its significance in linguistics. Information is provided on scientific research and solutions to problems between linguistics and literature through a new approach. The description of the Uzbek-Russian and Russian-Uzbek parallel corpus created on the basis of Abdulla Qodiriy’s novel “O‘tkan kunlar” is highlighted. The linguistic possibilities of the work, expressed in both corpora, and the peculiarities of the author’s pen are shown.

Аннотация. В данной статье рассматривается параллельный корпус, являющийся частью корпусной лингвистики, и его значение в лингвистике. Предоставляется информация о научных исследованиях и решении проблем между лингвистикой и литературой посредством нового подхода. Освещается описание узбекско-русского и русско-узбекского параллельного корпуса, созданного по роману Абдуллы Кадыри “Минувшие дни” Указываются языковые возможности произведения, выраженные в обоих корпусах, особенности, присущие перу автора.

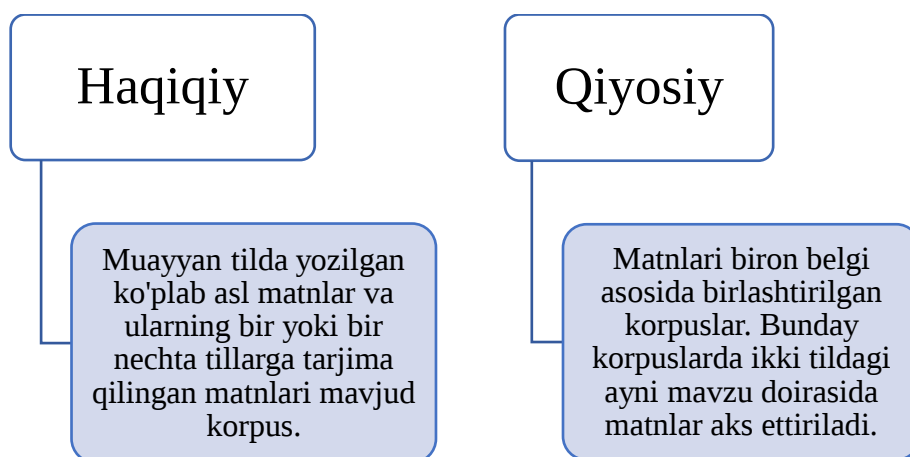
Kalit so‘zlar: Korpus lingvistikasi, korpus, parallel korpus, tezaurus, aligner tizimi, n-grammlar, konkordans.

Keyingi yillarda lingvistik tadqiqotlar sohasida yangicha yondashuvlar hamda metodologiyalarning paydo bo‘lishi tilshunoslikni yangi bosqichga olib chiqdi. Bu yangiliklardan biri parallel korpusning yaratilishi va uning lingvistik izlanishlarda keng foydalanishning yo‘lga qo‘yilishi bo‘ldi. **Parallel korpus** – biror tilning matnlarini boshqa tilga tarjima qilingan matnlar bilan o‘z ichiga olgan maxsus

ma'lumotlar bazasi sanaladi. Bunday korpus turi tilshunoslarga nafaqat tarjima jarayonini tahlil qilish, balki tilning leksik, sintaktik, morfologik va semantik xususiyatlarini chuqur o'rganish imkonini yaratadi.

Parallel korpuslar yoxud **tarjimalar korpusi** (ingl. *parallel / translation corpus*) – bu asl tildagi matnlar va uning boshqa tildagi tarjimasi mavjud korpus. Qoidaga ko'ra, bunday korpuslar faqat ikkita tilda tayyorlanadi, shu bois undagi o'zaro mutanosib matnlar *bitekst* ham deyiladi. **Bimatn (bitekst)** – asliy til matni va uning boshqa til(lar)dagi tarjimasi mavjud matnlar [1].

Parallellik nizomiga ko'ra korpuslar ikki asosiyga turga bo'linadi:



Har ikki parallel korpus ham lingvistik tadqiqotlarda muhim ahamiyat kasb etib bormoqda. Jumladan, tarjima tadqiqotlari, leksik tahlil, sintaktik tahlil, stilistik tahlil borasida til imkoniyatlari va tarjimonlarning tarjima usullarini aniqlash, ikki til borasidagi qiyosiy tahlillarni amalga oshirish hamda taqqoslashga yordam beradi. Parallel korpus yaratish jarayonida ba'zi qiyinchiliklar mavjud. Birinchi navbatda, yuqori sifatli tarjimalar topish va ularni tahlil qilishda yuzaga keladigan muammolarni keltirib chiqaradi. Ikkinchidan, ko'p tilli korpuslarni yaratish tildagi matnlarning o'zaro muvofiqlashgan(alignment) shaklini talab qiladi. Shunga qaramay, Parallel korpusning bir qator afzalliklarini sanab o'tamiz. Masalan, zamonaviy va keng qamrovli yondashuvlarni amalga oshiradi. Bu esa bir necha tillar yuzasidan keng qamrovli tahlil olib borish mumkinligini anglatadi. U tilni o'rganish va tarjimini tahlil qilishda qo'llaniladi. Tarjimaning to'g'riligi va mazmuni parallel matnlar yordamida baholanishi tarjima sifatini belgilashda qo'l keladi. Bunday korpuslar yordamida tilning semantik tizimi va so'zlarning kontekstual ma'nolarini o'rganish mumkin.

Badiiy asar bo'yicha parallel matnlar korpusi, avvalo, ikki yoki bir necha tillar grammatikasidagi o'xshash va farqli jihatlarni tahlil qilish imkonini beradi. Parallel matnlar korpusi har ikki tildagi badiiy uslub xususiyatlarini qiyoslashda

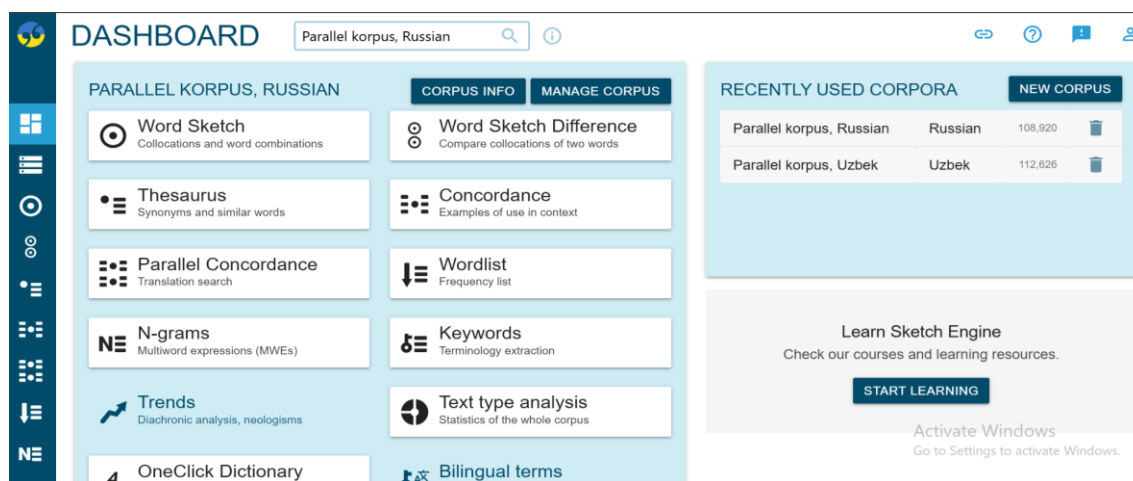
ahamiyatlidir [2]. Parallel korpuslar badiiy asarlarning original va tarjima qilingan variantlarini birlashtirib, tarjimadagi o'zgarishlarni, til va madaniyatdagi farqlarni, leksik va sintaktik strukturalarni solishtirishga imkon beradi.

Quyida yaratilgan Abdulla Qodiriyning “O'tkan kunlar” romani va uning rus tiliga tarjima qilingan nusxasi asosidagi parallel korpus ham tilshunoslik, tarjimashunoslik va adabiyotshunoslik uchun muhim ahamiyatga ega. Parallel korpus tuzishda asarning o'zbek tilida 2007-yilda nashrdan chiqqan nusxasi tanlab olindi. Rus-o'zbek parallel korpusi uchun Muhammadnodira Safarova tomonidan “Минувшие дни” nomi ostida 2009-yilda nashr qilingan nusxasi tanlab olindi. Korpusni yaratishda “SketchEngine” dasturidan foydalanilgan. Korpus haqidagi statistik ma'lumotlarni quyidagi jadvallarda ko'rish mumkin.

1-jadval. O'zbek-rus hamda Rus-o'zbek parallel korpusi ma'lumotlari

№	O'zbek-rus PK	Rus-O'zbek PK
Words	112,626	108,920
Tokens	158,194	140,759
Sentences	8,651	7175
Paragraphs	728	782

Ushbu jadvalda bir asarning ikki tildagi nusxalaridagi farqlarni yaqqol ko'rish mumkin. Jumladan, parallel korpusda asar rus tiliga tarjima qilinganda o'zbek tilidagi nusxasida mavjud so'zlarning 4000 tadan ko'proqqa qisqarganini ko'rish mumkin. Yoki tarjima nusxada gaplarning qariyb 1500 taga qisqargani ko'zga tashlanadi. Parallel korpus tuzishdan avval asarning ikki tildagi ayni nusxalarining elektron formatlari topib olindi. Keyin dasturga muvaffaqiyatli yuklanib, ikki tilli parallel korpus ishga tushdi (1-rasm).



1-rasm. Parallel korpus interfeysi

Asar tahliliga kirishishdan avval kalit soʻzlarni topib olish kerak boʻladi. Quyidagi jadvalda ikki korpusda topilgan dastlabki 15 ta kalit soʻzlarni koʻrish mumkin (2-jadval).

2-jadval. Parallel korpusdagi kalit soʻzlarning qiyosiy tahlili

Oʻzbek-rus PK			Rus-oʻzbek PK		
Kalit soʻzlar	Takrori	Baholash	Kalit soʻzlar	Takrori	Baholash
hasanali	189	1148.97	атабек	650	4555.61
qutidor	202	935.314	кумуш	554	3922.5
yusufbek	139	684.623	Kutubxonasi	430	3055.87
yonig	107	626.397	хасанали	276	1961.45
andek	115	626.039	зайнаб	283	1851.61
ach	110	578.605	атабека	258	1816.63
qolg	119	535.356	узбек-аим	200	1421.87
homid	155	528.829	офтоб-аим	176	1251.14
toshkand	113	523.543	кутидор	172	1222.95
chiqq	86	523.337	кушбеги	120	834.737
andan	75	475.101	хамид	137	804.019
ketkan	70	395.237	курбаши	91	631.597
tkan	282	389.307	маргилан	85	598.323
anlig	58	367.638	кутидора	78	555.139
mundan	67	365.094	Алим	87	509.835

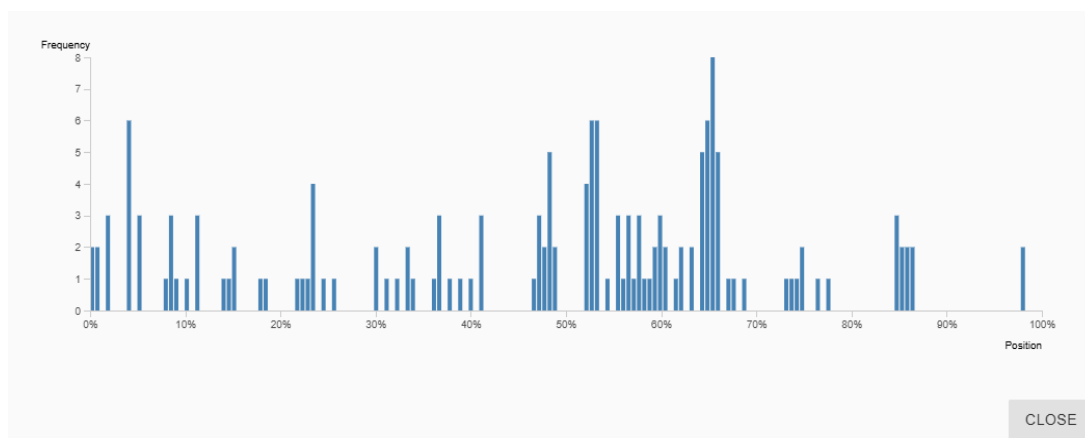
Ushbu jadvalni taqqoslaganda oʻzbek-rus va rus-oʻzbek parallel korpusidagi kalit soʻzlar tarjimada oʻzgarganini koʻrish mumkin. Jumladan, asl tilda “*Qutidor*”, “*Xasanali*”, “*Homid*”, “*Toshkand*” kabi soʻzlar eng koʻp qoʻllanilgan boʻlsa, rus tiliga tarjimada “*Атабек*”, “*Кумуш*”, “*Хасанали*”, “*Зайнаб*” kabi shaxs nomlarida koʻp takrordan foydalanilgan. Ushbu qiyosiy tahlildan yana bir xulosaga kelish mumkin boʻladi. Asarning tarjima nusxasida kishilarning lavozim, mansab hamda laqablari oʻzbek tilidan toʻgʻridan-toʻgʻri olib kirilgan. Tarjimon oʻzbek xalqi madaniyati va tarixiga oid ushbu soʻzlar ustida tarjimaga qoʻl urmagan.

Parallel korpusda badiiy asar tahliliga oid muhim jihat prallel konkordansning mavjudligidir. Har ikki asarning gaplar muvofiqligi-yu, ikki tilga xos boʻlgan leksik tuslanishlar ushbu tahlil jarayonida yuzaga chiqadi. Bunda parallel korpus kalit soʻzlarida yuzaga chiqqan soʻzlarni tanlasak, tahlil yana-da chuqurroq amalga oshiriladi. Masalan, “*Otabek*” soʻzini qidirdik. Bunda biz konkordans lugʻatda uchta tahlilni amalga oshirdik. Birinchidan, matnlar korpusda toʻgʻri muvofiqlashtirilganligi namoyon boʻldi (2-rasm).

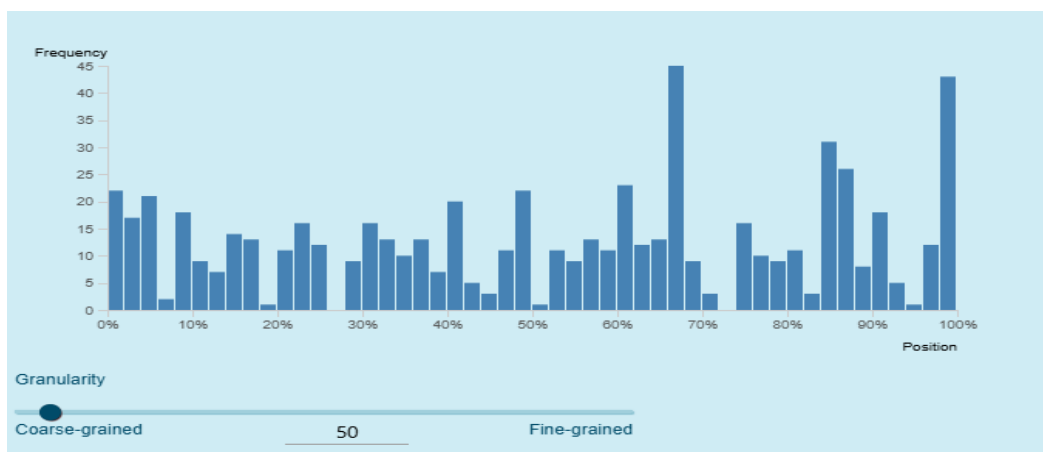
① docx0 <> Arabek ugochil g'ostey chae. </>	<> Otabek mehmonlarga choy quyib uzatdi. </>
① docx0 <> Imeet xonakomov so'ngimmi imenitimi ludiyli Tashkenta, v chisle koto rix, veroyatno, on xnal i vaxsho otca. - Vse vozmojno, - otvetil Arabek i pochemu-to nevolno vdragnul. </>	<> O'zi davlatmand bir kishi; Toshkand ashroflarining ko'blari bilan aloqador bo'lg'anligidan balki otangiz bilan tanish chiqar. </><> — Ehtimol, — dedi Otabek va nima uchundir g'ayri-ixtiyoriy bir tebrandi. </>
① docx0 <> Otca otca xelaxet pooboxat'sya s vami i iz vaxsh ust uslyashat'sya o delax v Tashkente. - Mne ne xotelos' bi pereexzhat iz otgo karavan-saraya, - otvetil Arabek, - a vot posetit' vaxsho otca ya gotov v lyuboe vrema. - Spasibo, bek-aka, ne naznachite li vse xhe den' vaxsho vizita? - Ya svobodn, kak vam izvestno, po vecheram, no esli vam udobnee drugoe vrema, y menya ne budet inoy vozmojnoxi, kak soglasitsya. - Budite v drazvni, - molvil Rahmat, - po zvol'te eche sprosit' vas: mojno </>	<> Otam sizning bilan o'turishib Toshkand ahvolotini so'zlash-makka mustoqdirar. </><> — Bu saroydan sizlarnikiga ko'chishim og'ir, — dedi Otabek, — ammo otangizning ziyoratlariga borishg'a har qachon hozirman. </>
① docx0 <> Odnako Arabek vse ne xaxodil. </>	<> Ammo Otabek nima uchundir tez kira bermadi. </>
① docx0 <> Posle etix slov Arabek vxoel v xudxru i pri sel y posteli. </>	<> Otabek bu so'z bilan hujraga kirdi va borib to'shagi yonig'a o'turdi. </>
① docx0 <> Odnako, Arabek kak sel y poste li, tak i xastyl v xadumxivosti... </>	<> Biroq Otabek to'shagi yonig'a o'turidda, yana o'ylab qoldi... </>
① docx0 <> Odnako, chem dolxise Arabek prebival v plenu xadumxivosti, tem bolxise neteprenie provil lya Hasanalni. - Net li dla menya porucheni? </>	<> Lekin Otabek xayol surishdan zerikmasa ham Hasanali kuzatishidan zerikdi. </><> — Menda yumushingiz yo'qmi? </>
① docx0 <> Arabek, kazalos', sluxayet i ne ponimayet, ibo, podnyv glaza, on vxlgnul na Hasanalni i vnoxe ustalils v odnu tochu. </>	<> Otabek bu gapni onglamadi shekillik, Hasanalining yuziga ko'tarilib qaradi-da, yana ko'zini bir nuqtag'a tikib qoldi. </>
① docx0 <> Arabek, naqone, slovo on nuvxish ot o na, vdragnul i vxlgnul na svoego slugu: - Otcho vy ne loxites' - Net li dla menya porucheni? - Kaxie moxut bit' porucheni, moxete idti i svexu toxo unexite. </>	<> Otabek shu holda tag'in qanchag'achadi o'turdi, nihoyat uyqudan uyg'ongan kishidek so'chib o'zini kuzatkuhiga qaradi: — Chiqib yotmayxizmi? </>
① docx0 <> Arabek vo vlasti xusti, Hasanalni ukradko poxlyaduyet na nego. </>	<> Otabek xomush o'turub, Hasanali ersa uning holini ta'qib etmakda edi. </>
① docx0 <> Arabek posmotrel na nego i utverditelno xiv nul. - Vot ux ne xaxo, - xazal on i, nemnogo po molxav, poxnil, - i vy xazetili moe unixie, pravwww.zyouz.com kutubxonasi </>	<> Otabek savol tashlag'uchig'a qarab oldi va tasdiq ishorasini berib: — Bilmadim, — dedi va bir oz to'xtab izoh berdi. </>
① docx0 <> Arabek bil izumen neoxidanim voprosom: - Ne moxu poniat' vas, otca, i vse xhe otvexu: po sei den' vy bili otcom ne toxlo dla menya, no i dla vsei naexei xemli i vsexda nam xelali lix' dobra. </>	<> Otabek, Hasanalining maqsadig'a tushunolmay maj-hul unga nazar tashladi: — Sizmi? </><> — deb kulimsiradi. </>
① docx0 <> Arabek vxryl pisyemo - ono bilo na bolxshom liste buma i pisanu tolxstym kalamom1. </>	<> Otabek xatni ochdi: katta qog'ozda, yo'g'on qalam bilan yozilg'an uzun bir maktub edi. </>
① docx0 <> Proxital pisyemo, Arabek vnoxe probexal vxlz dom po ne xvsem poniatnim mestam, a dostignuv mesta, gde izlagalos' ob Azixbeku, nevolno vxslil nul: "Ty i lical! </>	<> Otabek maktubni tugatib, tushunmay o'tkan jumalarni qayta ko'zdan kechirar ekan, Azixbek voq'asi ustiga kelganda ixtiyorsiz «tulki» deb yubordi. </>
① docx0 <> No, daxe vxnyxine vse, on vxjixdal, ego uderxivali taxie soobrazheniya: " Arabek molod, a v molodosti lybov' - xrylatay tixca. </>	<> Ammo bu to'g'rida Otabekning ko'magiga chinlab bel bog'lashdan uni ba'zi ehtimollar to'xtatib: « Otabek yosh, yoshlar muhabbati uchar qush, balki, bu kun-erta unutib ham yuborar» kabi mulohazalarda bo'ling'an edi. </>

2-rasm. Matnlar muvofiqlashgan(alignment) parallel korpus

Ikkinchi qiyosiy tahlil “Otabek” shaxs nomining ikki korpusda ham asarda qo‘llanilganlik darajasini diagrammada ifodalanishi bo‘ldi (3-rasm). Ikki korpus diagrammasida ham ko‘rish mumkinki, asarning 70% qismiga yaqinlashganda har ikki asarda ham “Otabek” ismidan eng yuqori darajada foydalanilgan. Va, aksincha, o‘zbek-rus parallel korpusida ushbu ismga asarning 85 foizidan boshlab oxirigacha ko‘p takrorda murojaat qilingan bo‘lsa (4-rasm), rus-o‘zbek parallel korpusida tarjima asarning 85% dan to oxirigacha bo‘lgan qismda umuman tilga olinmagan.

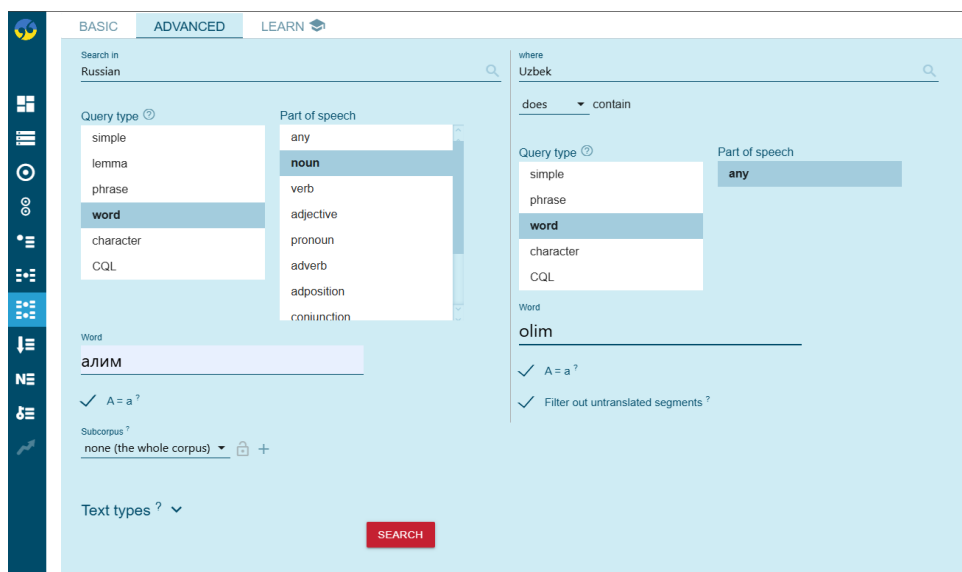


3-rasm. Rus tili parallel korpusida “Amaбек” shaxs nomining qo‘llanilishi



4-rasm. O'zbek tili parallel korpusda "Otabek" shaxs nomining qo'llanilishi

Keyingi qiyosiy tahlil parallel korpusda rus tili va o'zbek tilidagi so'zlarni qidirish imkoniyatlarini ko'rsatib berdi. Masalan, har ikki tili korpusda ham qidiruvni yakka, fraza, so'z va hokazo shakllar ustida amalga oshirish mumkinligi ko'rinadi. Biroq keyingi bosqichlarda rus tili parallel korpusining imkoniyatlari kengayadi. Ya'ni qidiruvdagi birlikning so'z turkumini, o'rnini aniqlash mumkin bo'ladi. O'zbek tilidagi asar korpusida esa so'z turkumlarini belgilash imkoni mavjud emasligi ayon bo'ldi (5-rasm).



5-rasm. Parallel konkordans qidiruv imkoniyatlari

Keyingi tahlilni amalga oshirish ham parallel korpus imkoniyatlarini ko'rsatib beruvchi muhim jihatdir. Bu N-grammlar asosida matnda qo'llanilgan turli leksik birliklarning bog'lanish shakllarini topish imkoni mavjud. Badiiy asarning lingvistik mohiyati ham ushbu tahlilda yuzaga chiqadi. Masalan, "O'tkan kunlar" asarining asl tilida 2-grammlardan foydalanib ko'rdik. Bunda eng ko'p qo'llanilgan so'z juftliklari, asosan, "O'zbekoyim", "Oftoboyim", "dedi va", "shuning uchun"

kabilar bo'lsa, rus tilidagi tarjima nusxasida “об этом”, “с ним”, “что бы” kabi birliklar ko'p qo'llanilgan.

3-jadval. 2-gramm asosida qiyosiy tahlil

	O'zbek tili PK	Rus tili PK
2-gramm natijasi	1037 ta	1367 ta
Takrori	12 690 marta	14 613 marta

Demak, Abdulla Qodiriyning “O'tkan kunlar” romani yuzasidan yaratilgan o'zbek-rus parallel korpusida statistik, qiyosiy, lingvistik tahlillarni amalga oshirib, quyidagi xulosalarni qayd etish mumkin bo'ldi:

1. Abdulla Qodiriyning “O'tkan kunlar” romani rus tilidagi tarjimasi korpusga joylanganda so'zlar soni asliyat tildagisidan 4000 taga, gaplar miqdori esa qariyb 1000 taga kamaygan. Bunda tarjima asarda ba'zi gaplar va jumlar qisqartirilgan.

2. Asar o'zbek tilidan rus tiliga tarjima qilinganda kishilarning lavozimi, mansab hamda laqabini bildiruvchi leksik birliklar to'g'ridan to'g'ri o'zbek tilidan rus tiliga o'tgan. Bunda asliyat tilga hamda o'zbek millati tarixiga xos birliklar saqlanib qolgan.

3. Asarning asliyat tilida 2-grammlar tahlili asosida shu aniq bo'ldiki, kishilarning nomlaridan keyingi o'rinda keng qo'llanilgan leksik birliklar qo'shma olmoshlar, ko'makchili birikmalardir.

4. Shaxs nomlari tarjima asarda o'zbek tilidagisidan 3 barobar ko'p qo'llanilgan (masalan, “Атабек” so'zi 650 marta). Bu esa tarjimonga xos xususiyatlardan bir ekanligi namoyon bo'lgan.

5. O'zbek-rus parallel korpusida parallel konkordans, tezaurus funksiyalari shuni ko'rsatdiki, rus tilidagi matnlar tahlilida korpus imkoniyatlari kengayadi. Buni o'zbek tilida korpus lingvistikasi hamda parallel korpus yangi soha ekani, hali taraqqiy etmagani hamda “Sketch Engine” dasturida ko'p tilli korpus yaratish imkoniyatlarining samaradorligini oshirish zarurligi bilan izohlaymiz.

Foydalanilgan adabiyotlar:

1. Abjalova M. Parallel korpuslarni yaratish asoslari / Kompyuter lingvistikasining zamonaviy texnologiyalari-CTCL 2023. –Tashkent, 2023. 35-42-b.
2. Akramdjanova M. Parallel matnlar korpusini yaratishning amaliy va nazariy ahamiyati / Ilm-fan va innovatsiya ilmiy-amaliy konferensiyasi (in-academy.uz/index.php/si). –Tashkent, 2023.
3. Mengliyev B., Hamroyeva Sh. Korpus lingvistikasi: Korpus tuzish va undan foydalanish. -Tashkent, 2020.



4. Sobirova N., Akramdjanova M., Sobirov Sh. Parallel korpus tuzish xususiyatlari, tuzilishi, interfeysi / Academic research in Educational Sciences, 2022. 34-b.
5. Abrurakhmonova N., Urdishev K. Corpus based teaching Uzbek as a foreign language. Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL), 2019.
6. https://app.sketchengine.eu/#dashboard?corpname=user%2Fzukhramarifbaeva%2Fparallel_korpus_russian