



POST-TAHRIRLANGAN VA INSON TARJIMASINING KORPUS ASOSIDAGI QIYOSIY TAHLILI

Haitova Maqsuda Kamol qizi

Buxoro shahridagi 43 sonli umumiy o'rta ta'lim maktabi

ingliz tili fani o'qituvchisi

Annotatsiya. Maqolada rus tilidan o'zbek tiliga o'girilgan mashina tarjimasini post-tahrirlash mahsuli tarjima universalialari nuqtayi nazaridan o'rganiladi. Material sifatida uch funksional uslubga oid kamida 400 segmentli uch qismli korpus loyihalashtirildi. Tadqiqotda korpus tahlili va parametrik bo'lmagan statistika qo'llanadi. Natijada to'rt universaliani o'lchaydigan takrorlanuvchi protokol ishlab chiqildi.

Kalit so'zlar. *post-tahrirlash, tarjima universalialari, korpus tarjimashunosligi, mashina tarjimasi, eksplitsitatsiya, simplifikatsiya, interferensiya.*

Kirish. O'zbekistonda tarjima buyurtmalarining salmoqli qismi endi tayyor mashina varianti bilan boshlanadi. Tarjimon matnni noldan o'girmaydi, balki mashina chiqarib bergan matnni tahrirlaydi. Post-tahrirlash talablari ISO 18587:2017 xalqaro standartida belgilangan [17]. Korpus tarjimashunosligida tarjima matnlari manba tildan qat'i nazar takrorlanadigan belgilarga ega degan qarash shakllangan. M.Beyker bunday belgilarni **tarjima universalialari** deb atagan va ularning to'rt turini, ya'ni eksplitsitatsiya, simplifikatsiya, normalizatsiya va tenglashtirishni ajratgan [1; 233-250-b.]. So'nggi o'n yillikda tadqiqotchilar bu belgilarni inson tarjimasida emas, mashina tarjimasi mahsulida va post-tahrirlangan matnda qidirmoqda. A.Toral 5 tarjima yo'nalishini qamrab olgan to'plamda post-tahrirlangan matn inson tarjimasiga



qaraganda soddaroq va ko'proq normallashtirganini, manba tilning ta'siri esa unda kuchliroq saqlanganini aniqlagan [8; 273-b.]. Bu natijalar ingliz, nemis, ispan va golland tillari uchun olingan.

Tadqiqot obyekti rus tilidan o'zbek tiliga o'girilgan uch turdagi matndan iborat bo'lib, ular xom mashina tarjimasi, post-tahrirlangan matn va mustaqil inson tarjimasidir. Tadqiqot predmeti shu uch guruhni bir-biridan ajratib turadigan universaliza belgilaridir. Maqsad post-tahrirlash mahsulini universalizatsiya nuqtayi nazaridan o'lchaydigan va boshqa tadqiqotchilar tomonidan takrorlanishi mumkin bo'lgan protokolni ishlab chiqishdan iborat. Maqsadga erishish uchun to'rt vazifa qo'yildi. Birinchi vazifa universalizatsiya nazariyasi va post-editsiya hodisasi bo'yicha ishlarni tanqidiy sharhlash hamda rus va o'zbek juftligiga nisbatan o'rganilmagan masalani aniqlashdan iborat. Ikkinchi vazifa o'zbek tilining agglutinativ tuzilishiga mos keladigan o'lchov ko'rsatkichlarini tanlash. Uchinchi vazifa uch guruhni qiyoslash uchun statistik tartib va qaror mezonlarini belgilash. To'rtinchi vazifa o'zbek tili uchun mavjud avtomatik tahlil vositalarining aniqligini baholash va qo'lda annotatsiyaga o'tish chegarasini asoslashdir.

Adabiyotlar tahlili va metodologiya. Mavzuga oid ishlarni yondashuvi bo'yicha to'rt guruhga ajratish mumkin. Birinchi guruh universalizatsiya tushunchasini shakllantirgan ishlardan iborat. M.Beyker tarjima matnini alohida til hodisasi sifatida o'rganishni taklif qilgan va korpus usullarini tarjimashunoslikka olib kirgan [1; 233-250-b.]. S.Laviosa simplifikatsiyani miqdoriy o'lchash uchun leksik zichlik, tip va token nisbati (TTR), o'rtacha gap uzunligi hamda yuqori chastotali leksika qamrovi kabi ko'rsatkichlarni asoslagan [2; 557-570-b.]. M.Olohan bilan M.Beyker ingliz tilidagi tarjimalarda ixtiyoriy bog'lovchi chastotasi tabiiy matnlardagidan yuqori ekanini ko'rsatib, eksplitsitatsiyaga empirik dalil keltirgan [3; 141-158-b.]. E.Chesterman universalizatsiyalarni ikkiga bo'lgan, chunki manba matn bilan



tarjimani qiyoslaganda ko'rinadigan belgilar tarjima matni bilan maqsad tilning tabiiy matnlarini qiyoslaganda ko'rinadigan belgilardan farq qiladi [4; 33-49-b.].

Ikkinchi guruh bu tushunchani tanqid qiladi. V.Bexer tarjimaga xos eksplitsitatsiya haqidagi qarashni rad etib, kuzatilayotgan farqlarni til juftligiga xos konvensional farqlar bilan izohlash kerakligini asoslagan [6; 1-28-b.]. Keyingi ishida u bog'lovchilarning qo'shilishini universal mayl emas, aniq juftlikdagi grammatik va uslubiy majburiyat natijasi deb talqin qilgan [7; 26-47-b.]. S.Bernardini bilan F.Zanettin universaliniyanning universaligi korpus tanloviga va qiyoslash bazasiga bog'liq ekanini ko'rsatgan [5; 51-62-b.]. Mazkur tanqid ushbu tadqiqotning maqomini belgilaydi. Quyida tekshiriladigan tasdiq barcha tillar uchun emas, faqat rus va o'zbek juftligi uchun ilgari suriladi.

Uchinchi guruh tadqiqot markazini inson tarjimasidan mashina mahsuliga ko'chirgan ishlarni qamrab oladi. A.Toral post-tahrirlangan matnni tarjimaning kuchaygan ko'rinishi sifatida tavsiflagan [8; 273-281-b.]. E.Vanmassenxove mashina tarjimasi tizimlari leksik xilma-xillikni kamaytirishini, chastotali qoliplarni kuchaytirib, kam uchraydiganlarini chetda qoldirishini ko'rsatgan [9; 222-232-b.]. J.Deyms sifat jihatidan teng matnlarda ham inson tarjimasi bilan post-tahrirlangan matnni ajratish mumkinligini aniqlagan [10; 89-103-b.]. M.Farrell post-tahrirlovchi maqbul yechimni odatda o'zgartirmasligini va shu sababli ayrim ifodalar post-tahrirlangan matnda inson tarjimasidagidan ko'ra tez-tez uchrashini kuzatgan hamda bunday birliklarni mashina tarjimasi markerlari deb atagan [11; 50-59-b.]. S.Kastilyo hamda N.Rezende post-tahrirlash belgilarini o'lchash uchun alohida resurs yaratgan [12; 305-306-b.].

To'rtinchi guruh post-tahrirlash jarayonini va uning mehnat hajmini o'lchaydi. H.Krings sa'y-harakatni uch o'lchovga, ya'ni vaqt, klavish bosishlar va aqliy yuklamaga ajratgan [13]. M.Koponen bu yo'nalishdagi tadqiqotlarni



umumlashtirib, tejalgan vaqt bilan yakuniy sifat orasidagi bog'liqlik bir xil emasligini ko'rsatgan [14; 131-148-b.]. M.Koponen, L.Salmi va M.Nikulin neyron tarmoq, statistik hamda qoidaga asoslangan tizimlar chiqishidagi tuzatishlarni qiyoslagan [15; 61-90-b.]. M.Snover tuzatish hajmini o'lchash uchun TER va HTER ko'rsatkichlarini taklif qilgan, bunda so'zlarni qo'shish, o'chirish, almashtirish va so'z ketma-ketligini siljitish amallari hisobga olinadi [16; 223-231-b.].

O'zbek tili uchun avtomatik tahlil vositalari yaqin yillarda ancha rivojlandi. U.Salayev qo'shimchalarga tayanadigan morfologik tahlilchini ishlab chiqqan [18], N.Abdurahmonova leksikon qidiruvi bilan affiks ajratishni birlashtirgan tizimni taqdim etgan [19], L.Bobojonova neyron tarmoq asosidagi so'z turkumi teglagichini yaratgan [20], A.Agostini esa o'zbek tili uchun leksik va semantik bazani tuzgan [21].

Material uch qismli birliklardan tuziladi. Har bir birlikda rus tilidagi manba matn, uning xom mashina tarjimasini va o'sha mashina variantining post-tahrirlangan ko'rinishi bo'ladi. Nazorat guruhini o'sha segmentlarning mashina variantiga murojaat qilinmasdan bajarilgan mustaqil inson tarjimasini tashkil etadi. Korpus hajmi kamida 400 segment yoki 50.000 so'zshakldan iborat bo'lib, badiiy, publitsistik hamda rasmiy va ilmiy uslublar teng ulushda qamrab olinadi. Matnlar ochiq litsenziyali nashrlardan olinadi, har bir manbaning nashri, muallifi, yili va litsenziyasi qayd etiladi. Takrorlanuvchi parchalar Jakkor o'xshashligi chegarasi bo'yicha filtrlanadi. Matnlarning mashina tarjimasini tizimlari o'quv ma'lumotlariga tushgan bo'lish ehtimoli nashr sanasi, ochiqlik holati va parcha qidiruvi orqali tekshiriladi hamda protokolda bayon qilinadi.

Natijalar. Uch funksional uslubni bir-biri bilan qiyoslash uchun har bir uslubda kamida 120 segment bo'lishi lozim, aks holda uslublararo farq



ishonchli tekshirilmaydi. Umumiy hajm kamida 400 segment yoki 50.000 soʻzshakl etib olindi.

1-jadval. Korpus tarkibi va hajm chegaralari

Uslub	Eng kam segment soni	Har bir segment uchun matn guruhlari
Badiiy	120	rus manba matni, xom mashina tarjimasi, post-tahrirlangan matn, mustaqil inson tarjimasi
Publitsistik	120	rus manba matni, xom mashina tarjimasi, post-tahrirlangan matn, mustaqil inson tarjimasi
Rasmiy va ilmiy	120	rus manba matni, xom mashina tarjimasi, post-tahrirlangan matn, mustaqil inson tarjimasi
Jami	kamida 400	kamida 50.000 soʻzshakl

Jadvaldan koʻrinadiki, har bir segment toʻrt matn guruhida qayd etiladi, shuning uchun yigʻiladigan matn hajmi segmentlar sonidan toʻrt baravar koʻp boʻladi. Pilot bosqichda bitta uslubda 40 tadan 60 tagacha segment va ikki post-tahrirlovchi bilan sinov oʻtkaziladi, bu esa toʻliq korpusning 10%dan 15%gacha qismini tashkil qiladi. Universalialar bevosita oʻlchanmaydi, ular koʻrsatkichlar orqali operatsionallashtiriladi. Har bir universalija uchun bir nechta koʻrsatkich tanlandi, chunki bitta koʻrsatkich tasodifiy ogʻishga sezgir boʻladi va yolgʻiz oʻzi ishonchli dalil bermaydi.

2-jadval. Universalialar va ularni oʻlchash koʻrsatkichlari



Universaliya	Ko'rsatkichlar	O'lchash birligi
Eksplitsitatsiya	bog'lovchilar zichligi, kirish konstruksiyalari zichligi, qo'shilgan izohlovchi elementlar soni, maqsad va manba matn uzunligining nisbati	1000 so'zshaklga nisbatan chastota va nisbat
Simplifikatsiya	leksik zichlik, TTR, MATTR, o'rtacha gap uzunligi, yuqori chastotali leksika ulushi	koeffitsiyent va foiz
Normalizatsiya	tipik kollokatsiyalarga siljish, tinish belgilarining normallashuvi, shablon iboralar chastotasi	chastota
Interferensiya	manba til so'z tartibining saqlanishi, kalkalar ulushi, rus sintaktik modellarining takrorlanishi	segmentlar ulushi, foiz

Jadvalda keltirilgan 15 ko'rsatkich 3 guruhda bir xil tartibda hisoblanadi. Shundan to'rttasi eksplitsitatsiyaga, beshtasi simplifikatsiyaga, uchtasi normalizatsiyaga va uchtasi interferensiyaga tegishli. Ko'rsatkichlar soni universalialar bo'yicha teng emas, chunki simplifikatsiya uchun korpus tarjimashunosligida allaqachon sinovdan o'tgan o'lchovlar mavjud [2; 557-570-b.], normalizatsiya va interferensiya esa ko'proq sifat tahliliga tayanadi. Tuzatish hajmi alohida blok sifatida hisoblanadi. U universalialar emas, biroq post-tahrirlovchi qanchalik aralashganini ko'rsatadi va universalialar ko'rsatkichlarini talqin qilishda zarur bo'ladi.

3-jadval. Tuzatish hajmi 63as a'y-harakatni o'lchash tartibi



Ko'rsatkich	Nima o'lchanadi	Hisoblash vositasi
TER	mashina variantidan post-tahrirlangan matnga o'tishdagi qo'shish, o'chirish, almashtirish va o'rin almashtirish amallari soni	tercom yoki sacrebleu
HTER	post-tahrirlovchi kiritgan haqiqiy tuzatishlar asosidagi tahrirlash darajasi	tercom yoki sacrebleu
Tuzatish turlari taqsimoti	to'rt amalning umumiy tuzatishlar ichidagi ulushi	tercom chiqishini qayta ishlaydigan skript
n-gramm mos kelishi	post-tahrirlashdan keyin saqlanib qolgan mashina matni ulushi	maxsus skript
Sa'y-harakat	segmentga sarflangan vaqt, klavish bosishlar soni, pauzalar taqsimoti	PET, Translog-II yoki Inputlog

Muhokama. Post-tahrirlovchi bo'sh sahifa bilan emas, tayyor matn bilan ishlaydi. Uning vazifasi matnni qayta yaratish emas, uni maqbul holatga keltirishdir. Mashina taklif qilgan yechim tushunarli va xatosiz bo'lsa, u odatda saqlanib qoladi. M.Farrell aynan shu holatni kuzatgan va post-tahrirlangan matnda ayrim ifodalar inson tarjimasidagidan ko'ra tez-tez uchrashini aniqlagan [11; 50-59-b.]. H.Krings modeli bu xulqni sa'y-harakat nuqtayi nazaridan izohlaydi, chunki har bir o'zgartirish qo'shimcha texnik hamda aqliy yuklama talab qiladi [13]. Rus va o'zbek tillarining sintaktik qoliplari bir xil emas. O'zbek tilida kesim gap oxirida keladi, ergash gap ko'pincha bosh gapdan oldin joylashadi, so'z tartibi esa gapning aktual bo'linishiga bo'ysunadi. Mashina tarjimasi manba matnning tartibini saqlashga moyil



bo'ladi. Natija tushunarli chiqsa, post-tahrirlovchi uni o'zgartirmaydi va rus sintaktik modelining izi matnda qolaveradi. Mustaqil tarjimon bunday chegaraga duch kelmaydi, u gapni o'zbek tilining odatdagi qolipida quradi. Interferensiya ko'rsatkichi shu farqni qayd etishga mo'ljallangan.

Agglyutinativ tuzilish yana bir omil kiritadi. Rus tilidagi bitta so'zshakl o'zbek tilida bir necha qo'shimchali shaklga aylanadi, ruscha ko'makchili birikma esa kelishik qo'shimchasi orqali beriladi. Mashina bunday almashinuvda eng chastotali variantni tanlaydi va shu tanlov matn bo'ylab takrorlanadi. Qo'shimchalarning xilma-xilligi kamayadi, morfologik boylik esa torayadi. Ushbu holat leksik zichlik va MATTR ko'rsatkichlarida ko'rinishi kutiladi. Eksplitsitatsiyaning maqomi post-tahrirlashda o'zgaradi. Mustaqil tarjimada tarjimon faqat manba matnga tayanadi, post-tahrirlashda esa uning oldida ikkita matn turadi, ya'ni manba matn va mashina varianti. Shu sababli E.Chesterman taklif qilgan ajratma bu holatda to'liq ishlamaydi [4; 33-49-b.]. Post-tahrirlangan matnni baholash uchun uchinchi qiyoslash nuqtasi, ya'ni xom mashina tarjimasi zarur bo'ladi.

Normalizatsiya ikki tomondan kuchayadi. Mashina eng ko'p uchraydigan birikmani tanlaydi, post-tahrirlovchi esa g'ayritabiiy kalkani o'zbek tilining odatdagi iborasiga almashtiradi. Ikkala harakat ham matnni tipik qoliplarga yaqinlashtiradi. Bunday mahsulot xatosiz bo'lishi mumkin, biroq unda mustaqil tarjimonga xos yechim xilma-xilligi kamayadi. V.Bexerning tanqidi bu yerda o'z kuchini saqlaydi, chunki kuzatilgan siljishni universal mayl deb emas, aniq juftlik va aniq ish tartibi natijasi deb talqin qilish lozim [6; 1-28-b.].

Xulosa. Rus tilidan o'zbek tiliga o'g'irishda post-tahrirlash mahsulini alohida matn turi sifatida o'lchash mumkin, buning uchun to'rt universaliyaga bog'langan 15 ko'rsatkich hamda uch guruhli qiyoslash yetarli. ISO 18587:2017 talablariga muvofiq to'liq post-tahrirlash yo'riqnomasiga ikkita



qo'shimcha tekshiruv kiritilishi maqsadga muvofiq [17]. Birinchisi leksik xilma-xillik ko'rsatkichi bo'yicha, ikkinchisi so'z tartibi bo'yicha bajariladi. Xatolar sonini sanash bu ikki belgini qayd etmaydi, chunki segment xatosiz holda ham mashina qolipini saqlab qolishi mumkin.

Adabiyotlar ro'yxati

1. Baker M. Corpus Linguistics and Translation Studies. Implications and Applications // Text and Technology. In Honour of John Sinclair / eds. M. Baker, G. Francis, E. Tognini-Bonelli. Amsterdam, Philadelphia: John Benjamins, – 1993. – P. 233-250.
2. Laviosa S. Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose // Meta. – 1998. – Vol. 43, – No. 4. – P. 557-570.
3. Olohan M., Baker M. Reporting that in Translated English. Evidence for Subconscious Processes of Explicitation // Across Languages and Cultures. – 2000. – Vol. 1, No. 2. – P. 141-158.
4. Chesterman A. Beyond the Particular // Translation Universals. Do They Exist? / eds. A. Mauranen, P. Kuusimäki. Amsterdam: John Benjamins, – 2004. – P. 33-49.
5. Bernardini S., Zanettin F. When Is a Universal Not a Universal? // Translation Universals. Do They Exist? / eds. A. Mauranen, P. Kuusimäki. Amsterdam: John Benjamins, – 2004. – P. 51-62.
6. Becher V. Abandoning the Notion of Translation-Inherent Explicitation. Against a Dogma of Translation Studies // Across Languages and Cultures. – 2010. Vol. 11, No. 1. – P. 1-28.



7. Becher V. When and Why Do Translators Add Connectives? // Target. – 2011. Vol. 23, No. 1. – P. 26-47.
8. Toral A. Post-editee. An Exacerbated Translationese // Proceedings of Machine Translation Summit XVII. Research Track. Dublin, – 2019. – P. 273-281.
9. Vanmassenhove E., Shterionov D., Way A. Lost in Translation. Loss and Decay of Linguistic Richness in Machine Translation // Proceedings of Machine Translation Summit XVII. Research Track. Dublin, – 2019. – P. 222-232.
10. Daems J., De Clercq O., Macken L. Translationese and Post-editee. How Comparable Is Comparable Quality? // Linguistica Antverpiensia, New Series. Themes in Translation Studies. – 2017. Vol. 16. – P. 89-103.
11. Farrell M. Machine Translation Markers in Post-Edited Machine Translation Output // Translating and the Computer 40. London: Asling, – 2018. – P. 50-59.
12. Castilho S., Resende N. MT-Pese. Machine Translation and Post-Editese // Proceedings of the 23rd Annual Conference of the European Association for Machine Translation. Ghent, – 2022. – P. 305-306.
13. Krings H. P. Repairing Texts. Empirical Investigations of Machine Translation Post-editing Processes / ed. G. S. Koby. Kent, Ohio: Kent State University Press, – 2001. – 635 p.
14. Koponen M. Is Machine Translation Post-editing Worth the Effort? A Survey of Research into Post-editing and Effort // The Journal of Specialised Translation. – 2016. No. 25. – P. 131-148.
15. Koponen M., Salmi L., Nikulin M. A Product and Process Analysis of Post-editor Corrections on Neural, Statistical and Rule-Based Machine



Translation Output // Machine Translation. – 2019. Vol. 33, No. 1. – P. 61-90.

16. Snover M., Dorr B., Schwartz R., Micciulla L., Makhoul J. A Study of Translation Edit Rate with Targeted Human Annotation // Proceedings of the 7th Conference of the Association for Machine Translation in the Americas. Cambridge, MA, – 2006. – P. 223-231.
17. Salaev U. UzMorphAnalyser. A Morphological Analysis Model for the Uzbek Language Using Inflectional Endings // AIP Conference Proceedings. – 2024. Vol. 3244, No. 1. Art. 030058.
19. Abdurakhmonova N., Ismailov A., Sayfulleyeva R. MorphUz. Morphological Analyzer for the Uzbek Language // 2022 International Conference on Computer Science and Engineering (UBMK). IEEE, – 2022.
20. Bobojonova L., Akhundjanova M., Ostheimer S., Fellenz S. BBPOS. BERT-based Part-of-Speech Tagging for Uzbek // Proceedings of the First Workshop on Language Models for Low-Resource Languages (LoResLM 2025). – 2025.
21. Agostini A., Usmanov T., Khamdamov U., Abdurakhmonova N., Mamasaidov M. UZWORDNET. A Lexical-Semantic Database for the Uzbek Language // Proceedings of the 11th Global Wordnet Conference. – 2021. – P. 8-19.
22. Kubedinova L., Khusainov A., Suleymanov D., Gilmullin R., Abdurakhmonova N. First Results of the TurkLang-7 Project. Creating Russian-Turkic Parallel Corpora and MT Systems // Proceedings of the International Conference on Computational and Corpus-Based Phraseology. CEUR Workshop Proceedings. – 2020. Vol. 2780.