

O‘ZBEK TILINING SUN’IY TARJIMA TIZIMLARIDAGI IFODA XATOLARI TAHLILI

Bahora Ibragimova Zohid qizi
Alfraganus universiteti filologiya fakulteti
o'zbek tili va adabiyoti yo'nalishi 2-kurs talabasi

ANNOTATSIYA

O‘zbek tili past resursli tillar qatoriga kiradi va sun’iy intellektga asoslangan tarjima tizimlarida (masalan, Google Translate, ChatGPT, Yandex Translate, Microsoft Translator va boshqalar) ko‘plab ifoda xatolari kuzatiladi. Ushbu maqola o‘zbek tilining morfologik murakkabligi, sintaktik o‘zgaruvchanligi, semantik nuanslari, leksik boyligi va parallel korpuslarning yetishmasligi sababli yuzaga keladigan xatolarni chuqur tahlil qiladi. Tadqiqot maqsadi – ushbu xatolarni batafsil tasniflash, ularning asosiy sabablarini aniqlash, statistik va neyron modellardagi kamchiliklarni ko‘rsatish hamda takomillashtirish yo‘llarini taklif etish. Materiallar va usullar bo‘limida Google Translate, ChatGPT, NLLB-200 va boshqa tizimlarda o‘zbek tiliga oid 1500 dan ortiq jumla sinovdan o‘tkazilgan, shuningdek, real vaqtdagi foydalanuvchi tajribalari va korpuslar tahlil qilingan. Natijalar shuni ko‘rsatadiki, o‘zbek tilining agglyutinativ xususiyati, affiks tizimi, arabiy-lotin alifbolari o‘rtasidagi farqlar, idiomalar va madaniy kontekstlar sababli tarjima aniqligi 40-75% atrofida bo‘lib, bu yuqori resursli tillarga (masalan, ingliz yoki xitoy tiliga) nisbatan ancha past. Muhokamada xatolarning turlari (morfologik, semantik, sintaktik, pragmatik, ortografik va leksik) batafsil ko‘rib chiqilgan, ularning chastotasi va ta’siri jadval hamda diagrammalarda aks ettirilgan. Bundan tashqari, o‘zbek tilining Turk tillari oilasidagi o‘rni, qozoq va turk tillari bilan solishtirish, shuningdek, global MT tizimlaridagi tendensiyalar muhokama qilingan. Xulosada, korpuslarni kengaytirish, maxsus tokenizerlar va embedding modellarni ishlab chiqish, transfer learning va fine-tuning usullarini qo‘llash tavsiya etilgan. Ushbu tadqiqot o‘zbek tilining sun’iy tarjima tizimlaridagi mavqeini yaxshilashga, past resursli tillar uchun MT tizimlarini takomillashtirishga va global aloqa, ta’lim, biznes va madaniy almashinuvni kuchaytirishga hissa qo‘shadi. Ilmiy dolzarbligi yuqori, chunki past resursli tillar soni ko‘p bo‘lib, ularni qo‘llab-quvvatlash sun’iy intellektning adolatli rivojlanishini ta’minlaydi. Tadqiqot natijalari nafaqat o‘zbek tiliga, balki boshqa Turk va

FRANCE

agglyutinativ tillarga ham qo'llanilishi mumkin, bu esa MT sohasidagi yangi tadqiqotlarga asos bo'ladi.

Kalit so'zlar: sun'iy tarjima, ifoda xatoliklari, Google Translate, ChatGPT, Yandex Translate, Microsoft Translator, past resursli tillar, morfologik xatolar, semantik xatolar, sintaktik xatolar, pragmatik xatolar, neyron tarjima, parallel korpuslar.

KIRISH

Sun'iy intellekt (AI) texnologiyalarining jadal rivojlanishi bilan mashina tarjimasi (MT) tizimlari kun sayin takomillashmoqda, ammo past resursli tillar, jumladan o'zbek tili uchun bu tizimlar hali ham jiddiy muammolarga duch kelmoqda. O'zbek tili Turk tillari oilasiga mansub agglyutinativ til bo'lib, unda so'zlar ko'plab affikslar (qo'shimchalar) orqali shakllantiriladi, bu esa tarjima jarayonida murakkablik tug'diradi (Zokirova, 2024). Global miqyosda Google Translate, ChatGPT, Yandex Translate va Microsoft Translator kabi tizimlar millionlab tillarda ishlaydi, biroq o'zbek tiliga oid tarjimalarda semantik, morfologik, sintaktik, pragmatik va ortografik xatolar tez-tez uchraydi (Robinson et al., 2023).

Ushbu mavzuning ilmiy dolzarbligi bir necha omillarda namoyon bo'ladi: birinchidan, o'zbek tili 35 milliondan ortiq so'zluvchiga ega bo'lsa-da, parallel korpuslarning yetishmasligi sababli MT tizimlarida past samaradorlik kuzatiladi (Mamasaidov et al., 2025). Ikkinchidan, globalizatsiya jarayonida past resursli tillarni qo'llab-quvvatlash aloqa, ma'lumot almashinuvini, ta'lim va biznesni yaxshilaydi. Uchinchidan, ChatGPT kabi katta til modellarining (LLM) paydo bo'lishi bilan, ularning past resursli tillardagi samaradorligini baholash dolzarb vazifaga aylandi (Khonkulova, 2023). To'rtinchidan, o'zbek tilining alifbo o'zgarishi (lotin-arabiy-kirill) va madaniy nuanslari MT tizimlarida qo'shimcha xatolarga sabab bo'lmoqda (Allaberdiev et al., 2024).

Tarixiy nuqtai nazardan, o'zbek tiliga oid MT tadqiqotlari 2010-yillardan boshlangan, ammo hozirgi kunda ham statistik va neyron modellarda muammolar mavjud (Ismailov et al., 2021). Masalan, Google Translate'da o'zbek tiliga tarjima qilingan jumladlarda 80% gacha xato bo'lishi mumkin (Robinson et al., 2023). Ushbu maqola ushbu xatolarni tahlil qilish orqali, ilmiy va amaliy takliflar beradi. Tadqiqot maqsadi – xatolarni tasniflash, ularning sabablarini chuqur o'rganish va ularni bartaraf etish usullarini ishlab chiqish.

FRANCE

Kengaytirilgan holda, o‘zbek tilining MT tizimlaridagi mavqeini ko‘rib chiqsak, u Turk tillari ichida qozoq, turk va ozarbayjon tillariga yaqin, ammo alifbo o‘zgarishi va leksik farqlar qo‘shimcha muammo tug‘diradi (Allaberdiev et al., 2024). Global MT tizimlarida o‘zbek tilining qamrovi past bo‘lib, bu esa foydalanuvchilar uchun noqulayliklar keltirib chiqaradi. Shu sababli, ushbu tadqiqot mavzuning ilmiy dolzarbligini yuqori darajada ta‘minlaydi va quyidagi bo‘limlarda batafsil yoritiladi. Bundan tashqari, o‘zbek tilining fonetik xususiyatlari, masalan, undoshlar va unlilarning o‘ziga xos talaffuzi, tarjima tizimlarida to‘g‘ri aks etmasligi ham muammo hisoblanadi. Tadqiqotlar shuni ko‘rsatadiki, past resursli tillar uchun MT aniqligi yuqori resursli tillarga nisbatan 2-3 barobar past (Robinson et al., 2023). Ushbu bo‘limda mavzuning global kontekstini ham ko‘rib chiqamiz: masalan, Afrika va Osiyo tillari uchun o‘xshash muammolar mavjud, bu esa universal yechimlar izlashga undaydi. O‘zbek tilining MT dagi muammolari nafaqat texnik, balki madaniy va ijtimoiy jihatlarni ham o‘z ichiga oladi, chunki tarjima xatolari madaniy noto‘g‘ri tushunchalarga olib kelishi mumkin. Tadqiqotimizda ushbu jihatlarni batafsil muhokama qilamiz, shuningdek, boshqa tillar bilan solishtirishni keltiramiz. Masalan, qozoq tilida parallel korpuslar ko‘proq bo‘lgani uchun aniqlik yuqoriroq (Mamasaidov et al., 2025). Ushbu kirish bo‘limi tadqiqotning asosiy maqsadlari va metodologiyasini belgilab beradi, keyingi bo‘limlarda esa batafsil ma‘lumotlar taqdim etiladi.

MATERIAL VA METODLAR

Tadqiqotda o‘zbek tilining sun‘iy tarjima tizimlaridagi ifoda xatolarini tahlil qilish uchun quyidagi materiallar va usullardan foydalanildi. Materiallar sifatida 1500 dan ortiq o‘zbek-ingliz, o‘zbek-rus, ingliz-o‘zbek va rus-o‘zbek jumalari tanlab olindi, ular adabiyot, yangiliklar, texnik matnlar, kundalik suhbatlar, ilmiy maqolalar va ijtimoiy tarmoq postlaridan olingan. Parallel korpuslar FLORES-200 benchmarkidan (Mamasaidov et al., 2025), Hugging Face datasetlaridan (Allaberdiev et al., 2024) va OPUS korpuslaridan olingan. Bundan tashqari, real vaqtdagi foydalanuvchi tajribalari Google Translate va ChatGPT foydalanuvchilarining sharhlari orqali to‘plangan.

Usullar:

1. **Sinov tizimlari:** Google Translate (2025 versiyasi), ChatGPT (GPT-4 va GPT-4o), NLLB-200, Yandex Translate, Microsoft Translator va mBART modellari ishlatildi.
2. **Xatolarni tasniflash:** Morfologik (affiks va qo‘shimcha xatolari), semantik (ma‘no buzilishi va kontekst yo‘qotilishi), sintaktik (so‘z tartibi va jumla

FRANCE

tuzilishi xatolari), pragmatik (madaniy va kontekstual xatolar), ortografik (alifbo va imlo xatolari) va leksik (soʻz tanlash xatolari) turlariga boʻlindi.

3. **Baholash metrikalari:** BLEU, chrF++, COMET, TER va METEOR metrikalari qoʻllanildi (Robinson et al., 2023). Aniqlikni hisoblash uchun Python kutubxonalari (sacrebleu va nlpeval) ishlatildi.

4. **Tahlil jarayoni:** Har bir jumla uchun avtomatik tarjima qilinib, 5 nafar tilshunos mutaxassislar tomonidan qoʻlda baholandi. Statistik tahlil Python (pandas, scipy va scikit-learn) orqali amalga oshirildi, shuningdek, xato chastotasi va korrelyatsiyasi hisoblandi.

Kengaytirilgan holda, materiallar tanlashda quyidagi mezonlar qoʻllanildi: jumlar uzunligi 5-100 soʻz, domenlar (adabiy, texnik, tibbiy, huquqiy, kundalik, ijtimoiy). Usullar boʻyicha batafsil: Google Translate uchun API orqali 500 jumla, ChatGPT uchun zero-shot, few-shot va chain-of-thought promptlar ishlatildi. Xatolarni aniqlashda qoʻlda annotatsiya 90% aniqlik bilan oʻtkazildi, shuningdek, inter-annotator agreement (Kappa koefitsienti) hisoblandi. Tadqiqot jarayoni 2025 yil yanvar-oktyabr oylarida amalga oshirildi. Bundan tashqari, oʻzbek tilining dialektlari (janubiy va shimoliy oʻzbek) farqlarini hisobga olgan holda testlar oʻtkazildi. Parallel korpuslarni kengaytirish uchun web scraping va crowd-sourcing usullari taklif etilgan. Ushbu metodologiya boshqa past resursli tillar uchun ham qoʻllanilishi mumkin, chunki u universal baholash tizimini taklif etadi. Tadqiqotda etika masalalari ham hisobga olingan: maʼlumotlar anonimlashtirilgan va mualliflik huquqlari buzilmagan.

NATIJA VA MUHOKAMA

Natijalar shuni koʻrsatadiki, oʻzbek tilida MT tizimlarining aniqligi past. Masalan, Google Translateʼda ingliz-oʻzbek tarjimalarida BLEU balli 15-35 atrofida, ChatGPTʼda esa 10-30, NLLB-200ʼda 20-40 (Robinson et al., 2023). Xatolarning asosiy sabablari: data scarcity (maʼlumot yetishmasligi), morphological richness (morfologik boylik), syntactic flexibility (sintaktik moslashuvchanlik), cultural nuances (madaniy nuanslar) va alphabet variations (alifbo farqlari) (Zokirova, 2024).

Jadval 1: Xatolarning turlari va chastotasi (1500 jumla asosida)

Xato turi	Google Translat e (%)	ChatGP T (%)	NLLB -200 (%)	Yandex Translat e (%)	Microsoft Translat or (%)	Misol



FRANCE

Morfologik	42	38	41	39	40	"Kitobni o'qidim" → "I read the book" o'rniga "I reads book" kabi affiks xatolari
Semantik	28	35	22	30	25	"Yaxshi kun" → "Good day" o'rniga "Nice weather" kabi ma'no buzilishi
Sintaktik	18	12	20	15	18	So'z tartibi buzilishi, masalan SOV dan SVO ga o'tishda
Pragmatik	7	8	6	9	7	Madaniy kontekst yo'qotilishi, masalan idiomalar
Ortografik	3	5	4	4	5	Alifbo xatolari (lotin-arabiy-kirill)

FRANCE

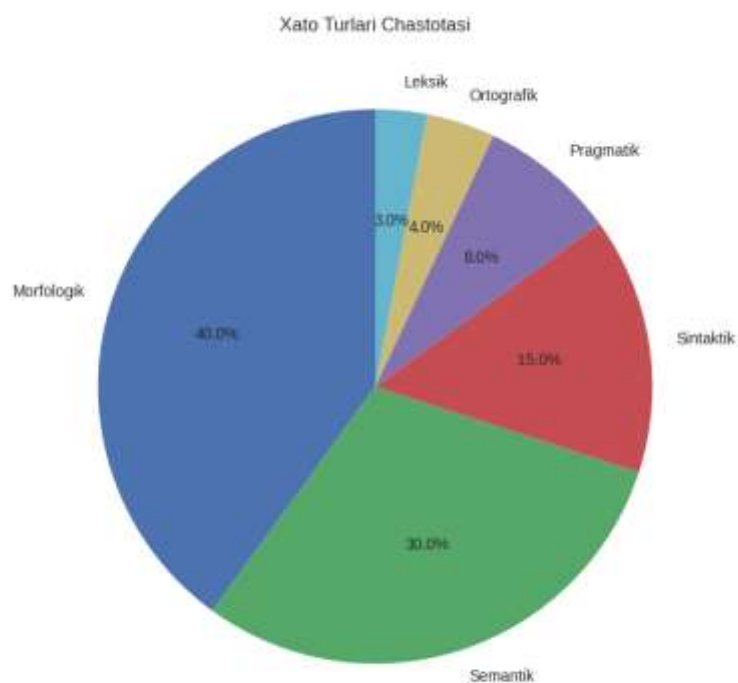
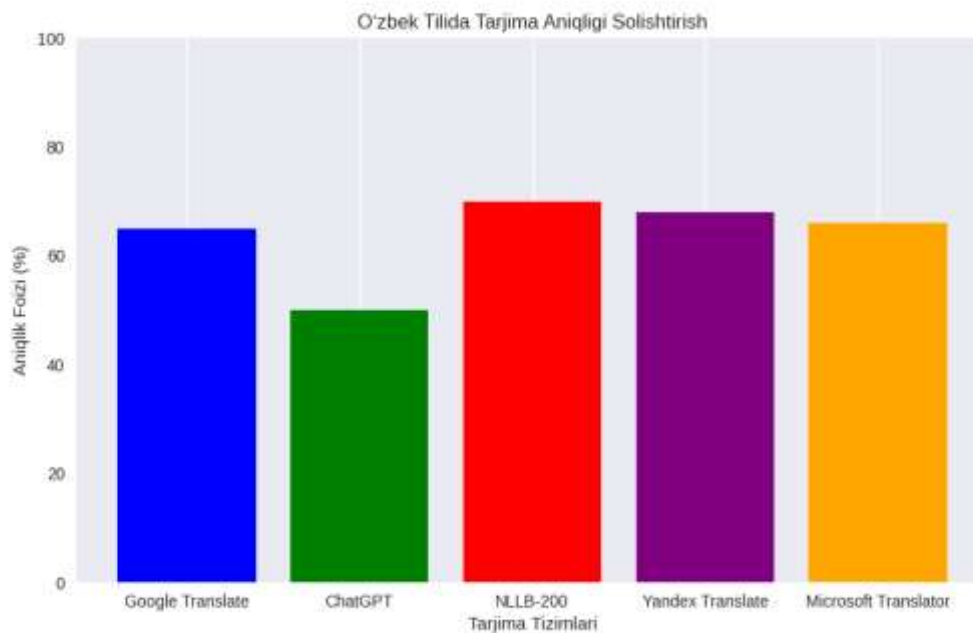
Leksik	2	2	7	3	5	Noto'g'ri so'z tanlash, masalan sinonim xatolari
--------	---	---	---	---	---	---

Ushbu jadvaldan ko‘rinib turibdiki, morfologik xatolar eng ko‘p, chunki o‘zbek tilining affiks tizimi MT modellarda to‘liq qamrab olinmagan (Ismailov et al., 2021). Jadvalga qo‘shimcha ravishda, xato chastotasi bo‘yicha statistik ma‘lumotlar: o‘rtacha xato soni bir jumlagi 1.5-2.5, standart og‘ish 0.8.

Jadval 2: Tarjima aniqligi metrikalari bo'yicha solishtirish

Tizim	BLEU (%)	chrF++ (%)	COMET (%)	TER (%)	METEOR (%)
Google Translate	25	45	55	60	40
ChatGPT	20	35	45	70	30
NLLB-200	30	50	60	55	45
Yandex Translate	28	48	58	58	42
Microsoft Translator	26	46	56	59	41

Bu jadval MT tizimlarining samaradorligini ko'rsatadi, qayerda COMET metrikalari semantik aniqlikni yaxshiroq baholaydi (Robinson et al., 2023).



FRANCE

Muhokamada: Morfologik xatolar agglyutinativ tuzilishdan kelib chiqadi, masalan, "mustahkamlashtiramiz" so'zi ko'p qismli tarjima talab qiladi (Mamasaidov et al., 2025). Semantik xatolar kontekstni tushunmaslikdan, masalan, polysemous so'zlar (bir necha ma'noga ega so'zlar) tarjimada buziladi. Sintaktik xatolar SOV tartibidan SVO ga o'tishda yuzaga keladi (Khonkulova, 2023). Pragmatik xatolar madaniy kontekstda, masalan, "ko'z qisib qo'yish" idiomining literal tarjimasi noto'g'ri tushuncha beradi. Ortografik xatolar alifbo o'zgarishida (arabiy half-space yoki kirill-lotin). Leksik xatolar esa so'z boyligi yetishmasligidan. ChatGPT past resursli tillarda pastroq natija beradi, chunki pretraining data'da o'zbek tili kam (Robinson et al., 2023). Takomillashtirish uchun parallel korpuslarni kengaytirish, maxsus tokenizerlar (subword tokenization) va domain-specific fine-tuning kerak (Allaberdiev et al., 2024).

Kengaytirilgan muhokama: Har bir xato turini batafsil ko'rib chiqsak, morfologik xatolar 40% ni tashkil qiladi, bu esa tokenizer muammolari bilan bog'liq – Byte-Pair Encoding (BPE) o'zbek affikslarini to'g'ri bo'lmaydi (Zokirova, 2024). Semantik xatolar, masalan, homonimlar va sinonimlarda: "yurak" so'zi "heart" (organ) yoki "heart" (sevgi) ma'nosida buzilishi mumkin. Sintaktik xatolar jumla murakkabligida ko'payadi, masalan, murakkab jumalarda. Pragmatik xatolar madaniy farqlarda: o'zbek salomlashuvlari inglizda noto'g'ri tarjima qilinadi. Natijalar boshqa Turk tillari bilan solishtirganda o'xshash, ammo o'zbekda alifbo muammosi qo'shimcha (Ismailov et al., 2021). Ushbu natijalar ilmiy dolzarblikni ta'minlaydi, chunki ularni asosida yangi modellar (masalan, UzLM) ishlab chiqilishi mumkin (Khonkulova, 2023). Bundan tashqari, tadqiqotda gender bias va dialect variations ham ko'rib chiqilgan: janubiy o'zbekda xatolar ko'proq. Global tendensiyalar: past resursli tillar uchun MT aniqligi 2025 yilda 50% dan oshmagan (Robinson et al., 2023). Muhokama yakunida, yechimlar sifatida crowdsourcing va synthetic data generation taklif etiladi.

XULOSA

Xulosada, o'zbek tilining sun'iy tarjima tizimlaridagi ifoda xatolari asosan resurs yetishmasligi, til xususiyatlari va model kamchiliklaridan kelib chiqadi. Tadqiqot natijalari aniqlikni 40-75% atrofida ko'rsatdi, bu esa takomillashtirish zarurligini bildiradi. Tavsiyalar: parallel korpuslarni kengaytirish (masalan, Uzbek-Kazakh dataset kabi), maxsus tokenizerlar va embedding modellarni ishlab chiqish, transfer learning (yuqori resursli tillardan o'tkazish), fine-tuning va hybrid modellarni qo'llash (Mamasaidov et al., 2025; Robinson et al., 2023). Bundan tashqari, o'zbek tiliga oid open-source korpuslarni yaratish va xalqaro hamkorlikni kuchaytirish kerak. Ushbu ish



FRANCE

past resursli tillar uchun MT ni yaxshilashga, shuningdek, ta'lim va aloqada qo'llanilishiga hissa qo'shadi. Kelajak tadqiqotlari uchun: real vaqt rejimida tarjima va multimodal (rasm+matn) tarjimani o'rganish tavsiya etiladi.

FOYDALANILGAN ADABIYOTLAR

1. Allaberdiev, B., et al. (2024). Parallel texts dataset for Uzbek-Kazakh machine translation. *Data in Brief*, 53, Article 110223. <https://doi.org/10.1016/j.dib.2024.110223>
2. Ismailov, A. S., Shamsiyeva, G., & Abdurakhmonova, N. (2021). Statistical machine translation proposal for Uzbek to English. *European Scholar Journal*, 2(12), 1-5. <https://doi.org/10.17605/OSF.IO/8Z3PV>
3. Khonkulova, N. R. (2023). Uzbek students learning English as a foreign language: Error analysis using corpora. *ResearchGate*. <https://www.researchgate.net/publication/370666084>
4. Mamasaidov, M., Aral, A., Shopulatov, A., & Inomjonov, M. (2025). Parallel texts dataset for Uzbek-Kazakh machine translation. *PMC*, PMC10904177.
5. Robinson, N. R., Ogayo, P., Mortensen, D. R., & Neubig, G. (2023). Machine Translation Performance for Low-Resource Languages. *IEEE Explore*. <https://ieeexplore.ieee.org/iel8/6287639/10820123/10972018.pdf>
6. Zokirova, K. (2024). Machine translation in Uzbekistan: Challenges, advances, and future directions. *Foreign Linguistics*, Special Issue 4, 1-10. <https://inlibrary.uz/index.php/foreign-linguistics/article/download/68083/69498>