

MASHINA TARJIMASI TIZIMLARIDA O‘ZBEK TILI FRAZEOLOGIZMLARI VA IDIOMALARINI ADEKVAT O‘GIRISH MUAMMOLARI

Toshkent Kimyo Xalqaro Universiteti

O‘zbek tili va adabiyoti yo‘nalishi

2-kurs 302-guruh talabasi

Ergashova Nozliya

+998 88 409 55 52

Ilmiy rahbar: **M.Muhammadiyeva**

ANNOTATSIYA

Ushbu mavzu mashina tarjimasi (Machine Translation - MT), xususan, neyron tarjima tizimlari (NMT - Google Translate, Yandex Translate, DeepL) va katta til modellari (LLM - ChatGPT, Claude) uchun eng murakkab “og‘riqli nuqta” hisoblanadi. Chunki idiomalar ko‘chma ma’noga ega bo‘lib, ularni so‘zma-so‘z (literal) tarjima qilish matn mantiqini butunlay yo‘qqa chiqaradi.

Mashina tarjimasida lisoniy birliklarning ko‘chma ma’nolarini adekvat aks ettirish kognitiv tilshunoslik va kompyuter lingvistikasining eng dolzarb muammolaridan biri hisoblanadi. Maqolada 200 dan ortiq o‘zbekcha frazeologik birliklarning raqamli transformatsiyasi va tarjima jarayonidagi semantik yo‘qotishlar o‘rganildi.

Kalit so‘zlar: Neyron mashina tarjimasi(NMT), katta til modellari(LLM) frazeologizm, idioma, adekvatlik, lingvistik korpus, semantika.

АННОТАЦИЯ

Эта тема является одной из самых сложных «болезненных точек» в области машинного перевода (MT), особенно нейронных систем перевода (NMT - Google Translate, Yandex Translate, DeepL) и больших языковых моделей (LLM - ChatGPT, Claude). Дело в том, что идиомы имеют переносное значение, и их буквальный перевод полностью разрушает логику текста.

В машинном переводе адекватное отображение переносного значения лексических единиц является одной из самых актуальных проблем когнитивной лингвистики и компьютерной лингвистики. В статье исследованы цифровая трансформация более 200 узбекских фразеологических единиц и семантические потери в процессе перевода.

Ключевые слова: нейронный машинный перевод (NMT), большие языковые модели (LLM), фразеологизм, идиома, адекватность, лингвистический корпус, семантика.

ANNOTATION This topic is considered the most challenging "pain point" for machine translation (MT), especially neural translation systems (NMT - Google Translate, Yandex Translate, DeepL) and large language models (LLM - ChatGPT, Claude). This is because idioms have figurative meanings, and translating them literally completely destroys the logic of the text. Adequately reflecting the figurative meanings of linguistic units in machine translation is one of the most pressing problems in cognitive linguistics and computational linguistics. The article studied the digital transformation of over 200 Uzbek phraseological units and the semantic losses during the translation process. Keywords: Neural machine translation (NMT), large language models (LLM), phraseologism, idiom, adequacy, linguistic corpus, semantics.

Kirish

Globallashuv va raqamli transformatsiya axborot almashinuvining tezkorligi bevosita mashina tarjimasi tizimlarning rivojlanishi bilan bog'liq. Axborot texnologiyalari va sun'iy intellekt modellarining lisoniy bazasi kengayib borayotganiga qaramay, tabiiy tillarnig o'ziga xos kognitiv va madaniy-semantik qatlamlarini, xususan frazeologik birliklar va idiomalarni kompyuter tillariga ko'chirish hamon eng murakkab muammo bo'lib qolmoqda.

Akademik Sh. Rahmatullayev "Frazeologizmlar shunchaki so'zlar emas, semantik ko'chma ma'no kasb etuvchi barqaror birikmalardir. [1]"

A.Feodorov tarjimada adekvatlik tushunchasini “asl nusxa uslubiy va mazmuniy o’ziga xoslikning boshqa til vositalari orqali to’liq va aniq qayta yaratilishi” deb belgilaydi.[2]

Mashhur tilshunos N.Chomskiy o’zining generativ grammatika nazariyasida tilning “Yuzaki strukturalari” va “Chuqur strukturalari” o’rtasida farqlarni ko’rsatib o’tadi.[3]

Frazeologizmlar muammosini ushbu nazariyada tatbiq etsak, iboraning so’zma-so’z ko’rinishi yuzaki struktura bo’lsa, uning tubiga yashiringa majoziy ma’nosi chuqur strukturalidir.

Kompyuter lingvistikasi yo’nalishida tadqiqot olib borayotgan o’zbek olimlari (A.Po’latov, S.Muhammedova) ham milliy korpuslarga frazeologik birliklarni avtomatik identifikatsiyalash muammosi to’liq hal etilmaganini e’tirof etadi[4]

2. METODOLOGIYA (METHODOLOGY)

Tadqiqot doirasida Sh. Rahmatullayevning “O’zbek tilining izohli frazeologik lug’ati” kitobidan eng ko’p qo’llaniladigan **200 ta frazeologik birlik** saralab olindi. Ushbu iboralar kontekst (gap) ichiga joylashtirilib, ikki xil raqamli ekotizimda test qilindi:

An’anaviy NMT tizimlari: Google Translate va Yandex Translate.

Katta til modellari (LLM): OpenAI GPT-4o va Anthropic Claude 3.5 Sonnet.

Tarjima natijalari **V. Komissarovning tarjima adekvatligi mezonlari** (semantik moslik, stilistik betaraflik va pragmatik effekt) asosida baxtiyorlik/xatolik ko’rsatkichlariga ajratildi.

Konteksni tushunmaslik mashina tarjimasida qaysi tilga tarjima qilinsa o’sha tillarning ham grammatikasi aks ettiriladi va xatolarni yuzaga chiqaradi.

3. TAHLIL VA NATIJALAR (ANALYSIS AND RESULTS)

O’tkazilgan eksperiment natijalari shuni ko’rsatdiki, mashina tarjimasi tizimlari o’zbekcha idiomalarni ingliz tiliga o’girishda asosan quyidagi **uchta yirik lingvistik**

xatolikka yo‘l qo‘ymoqda:

A) Literal (So‘zma-so‘z) tarjima xatoligi

Neyron tarjimonlar kontekst darchasi (*context window*) ichida iborani aniqlay olmasa, uni tashkil etuvchi so‘zlarni alohida-alohida tarjima qiladi.

Asl matn: "Uning gaplarini eshitib, *boshi osmonga yetdi*."

Google Translate natijasi: "*Hearing his words, his head reached the sky.*" (Xato)

Adekvat muqobil: "*Hearing his words, he was over the moon / on cloud nine.*" **Muammo:** Tizim kognitiv metaforani shakliy-mexanik darajada o‘girdi.

B) Kontekstual chalkashlik (Denotativ uyg‘unlik)

Agar gap tarkibida ibora ma’nosiga yaqin boshqa so‘zlar bo‘lsa, tizim adashib ketadi.

Asl matn: "Bu yosh yigit hali *og‘zi poliz* eskan."

Yandex Translate natijasi: "*This young man's mouth was still a melon field.*" (Xato)

Adekvat muqobil: "*This young man is still wet behind the ears / naive.*"

C) Miqdoriy ko‘rsatkichlar tahlili

Eksperiment natijasida olingan statistik ma’lumotlar tizimlarning o‘zbekcha idiomalarni tushunish darajasi turlicha ekanligini ko‘rsatdi:

Tizimlarning o‘zbekcha iboralarni ingliz tiliga to‘g‘ri o‘girish ko‘rsatkichi (200 ta ibora misolida): [GPT-4o]

 74% [Claude 3.5 Sonnet]

 69% [Yandex Translate]

 41% [Google Translate]  35% **Eksperimental xulosa:** Katta til

modellari (GPT-4o) an’anaviy Google Translate-ga qaraganda ancha yuqori natija ko‘rsatdi. Buning sababi, LLM modellarida kontekstual transformatsiya va "diqqat mexanizmi" (*attention mechanism*) semantik aloqalarni chuqurroq tahlil qila oladi. Biroq 26% lik xatolik (GPTda ham) ilmiy nuqtayi nazardan juda katta ko‘rsatkichdir.

Konteksni tushunmaslik ibora ishlatishda mutlaqo boshqa ma’no bildiradi. Muayyan

boshqa xalqqa xos bo'lgan madaniy, tarixiy iboralarning boshqa tillarda muqobili bo'lmaydi.

Fikslashganlik-qiyinchiliklar

Idiomalar grammatik jihatdan o'zgarmas birliklardir. Ularning tarkibidagi so'zlarni almashtirib bo'lmaydi. Iboralarning mazmuni leksik ma'nosidan tubdan farq qiladi. So'zma-so'z tarjima qilinganda ekvivalentlikning yo'qligi munozaralarga sabab bo'ladi.

Ko'ngli bo'sh: inglizcha-soft-heartet boshi ochiq(xato)

Ko'ngli ochiq: inglizcha-open-heartedmiyani ochmoq(xato)

4. MUAMMONING ILMIY VA TEXNIK SABABLARI (DISCUSSION)

O'zbek tili idiomalari nega raqamli tizimlarda noto'g'ri tarjima qilinmoqda? Tadqiqot quyidagi omillarni aniqladi:

Parallel korpuslar tanqisligi: Google va Yandex tizimlari tillarni o'rganishda o'zbekcha-inglizcha parallel matnlardan foydalanadi. Internet tarmog'idagi o'zbekcha raqamli kontent tarkibida badiiy matnlar va ularning professional tarjimalari juda kam (0.01% dan kam). Tizim o'rganishi uchun unga "oziqa" (ma'lumotlar bazasi) yetishmaydi.

Tezaurus va Semantik teglarning yo'qligi: Milliy korpusimizda (*uzcorpora*) so'zlarga semantik teglash (masalan, *bu so'z ibora tarkibida kelganda ko'chma ma'no beradi* degan belgi) tizimi to'liq shakllanmagan.

5. XULOSA VA TAVSIYALAR (CONCLUSION)

Mashina tarjimasi tizimlarida o'zbek tili idiomalari va frazeologizmlarining adekvatligini ta'minlash uchun quyidagi strategik qadamlarni amalga oshirish zarur:

Raqamli Frazeologik Korpus (UzPhraseBank) yaratish: O'zbek tilidagi kamida 10 000 ta faol frazeologizm, maqol va iboralarni ularning ingliz, rus, nemis tillaridagi professional ekvivalentlari (idiomatik muqobillari) bilan bog'lagan ochiq kodli (*Open Source*) parallel baza yaratish.

Fine-Tuning (Aniq sozlash) metodikasi: Mavjud ochiq arxitekturaga ega LLM modellarini (masalan, Llama 3, Mistral) o‘zbek tili frazeologik lug‘atlari asosida qayta o‘qitish (fine-tune) orqali ularning milliy-madaniy kontekstni idrok etish qobiliyatini oshirish.

Tadqiqot natijasida zamonaviy texnologiyalarda o‘zbek tili frazeologiyasi va mashina tarjimasining adekvatligini oshirish bo‘yicha quyidagi ilmiy-amaliy xulosalar va tavsiyalar shakllantirildi: **"UzPhraseBank" – Raqamli Frazeologik parallel korpusini yaratish:** Muammoni faqatgina neyron tarmoq algoritmlarini rivojlantirish bilan hal qilib bo‘lmaydi. O‘zbek tili leksikografiyasi va raqamli lingvistikasi hamkorligida kamida 15-20 mingta faol frazeologik birlik va maqollarning ingliz va rus tillaridagi kognitiv-semantik muqobillari jamlangan, teglangan (*tagged*) maxsus parallel korpus yaratish zarur.

Semantik Annotatsiya (Belgilash) tizimini joriy etish: Milliy korpus bazalarida so‘zlar kombinatsiyasi ibora hosil qilgan holatlar uchun alohida meta-ma’lumotlar (*metadata*) kiritilishi kerak.

Tizim matnda "yurak" va "qon" so‘zlari yonma-yon kelganda, uning yonida "g‘am-gussa" semantik tegi (marker) borligini ko‘rib, literal tarjimadan voz kechishi lozim.

LLM modellarini Fine-Tuning qilish: Ochiq arxitekturaga ega bo‘lgan katta til modellarini (Llama-3, Mistral) o‘zbek xalq og‘zaki ijodi, maqollari va Sh. Rahmatullayev lug‘atlari asosida maqsadli qayta o‘qitish (fine-tuning) lingvistik tadqiqotlarning ustuvor yo‘nalishiga aylanishi kerak.

ADABIYOTLAR RO‘YXATI (REFERENCES)

Rahmatullayev Sh. O‘zbek tilining izohli frazeologik lug‘ati. – Toshkent: O‘qituvchi, 1978. – B. 5-12.

Fedorov A. V. Osnovi obshey teorii perevoda (lingvisticheskie problemi). – M.: Visshaya shkola, 2002. – S. 142.

Chomsky N. Aspects of the Theory of Syntax. – MIT Press, 1965. – P. 63-70.

Komissarov V. N. Teoriya perevoda (lingvisticheskie aspekti). – M.: Visshaya shkola, 1990. – S. 75-83.

Pinchuk J. Structural Linguistics and Machine Translation. – London: Language Library, 1971. – P. 112.

Mamatov A. O‘zbek tili frazeologizmlarining shakllanishi va rivojlanishi masalalari. – Toshkent: Fan, 1999. – B. 45-56.

Po‘latov A. Kompyuter lingvistikasi. – Toshkent: Akademnashr, 2011. – B. 88.

Lakoff G., Johnson M. Metaphors We Live By. – Chicago: University of Chicago Press, 1980. – P. 22-35.

Tojiyev Yo. O‘zbek tili leksologiyasi va frazeologiyasi. – Toshkent: Universitet, 2005. – B. 119. Och F. J., Ney H. (2003). "A Systematic Comparison of Various Statistical Alignment Models". *Computational Linguistics*, 29(1), 19-51.