

RESEARCH ARTICLE

Problems And Solutions In Automatic Translation Of Uzbek Idioms And Figurative Expressions

Gulshoda Abdumalikova

Master's Student, Tashkent State University of Oriental Studies Tashkent, Uzbekistan

Charos Tukhtasinova

Master's Student, Tashkent State University of Oriental Studies Tashkent, Uzbekistan

VOLUME: Special Issue 2026

PAGE: 183-185

Copyright © 2026 European International Journal of Multidisciplinary Research and Management Studies, this is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-Share Alike 4.0 International License. Licensed under Creative Commons License a Creative Commons Attribution 4.0 International License.

Abstract

This article analyzes the challenges encountered in the automatic translation of Uzbek idioms and figurative expressions, examines the errors of NMT models, and proposes a computational linguistic model designed to identify idioms and translate them figuratively based on UzbekWordNet and semantic networks.

KEY WORDS

Machine translation, NMT, literal translation, idiom, WordNet, figurative expression.

INTRODUCTION

Tilning eng murakkab va o'ziga xos qatlamlaridan biri – bu idioma-lar, frazeologik birliklar va obrazli iboralaridir. Bu birliklar xalq-ning dunyoqarashi, madaniy tajribasi, mentaliteti hamda tarixiy xo-ti-rasini mujassam etgan ko'p qatlamli majoziy ma'nolarga ega bo'l-gan barqaror konstruksiyalardir. Shuning uchun ham idiomalar so'z-ma-so'z tarjima qilinmaydi, balki ular kontekstga, madaniy moslikka va semantik chuqurlikka tayanib tarjima qilinadi.

Bugungi globallashuv davrida tarjima jarayonining katta qismi av-to-matlashtirilmoqda. Mashina tarjimasi, ayniqsa neyron tarmoq-lar asosidagi NMT(neural machine translation) modellarining ri-voj-la-nishi natijasida ko'plab matnlarni tez va nisbatan sifatli tarjima qilish imkoniyati paydo bo'ldi. Biroq shunga qaramay, idiomatik birliklar mashina tarjimasi tizimlarida eng ko'p buziladigan va noto'g'ri tarjima qilinadigan segmentlar bo'lib qolmoqda. NMT modellarining literal tarjimaga moyilligi, madaniy kontekstni to'liq anglay olmasligi va semantik ko'p qatlamlilikni talqin qilishdagi cheklovlari idiomalar tarjimasida jiddiy xatolarga sabab bo'ladi. Shu

sababli, idiomalarni avtomatik aniqlash, semantik tahlil qilish va ularni to'g'ri tarjima qilishga mo'ljallangan maxsus kompyuter lingvistik modellarga ehtiyoj ortib bormoqda. Xususan, o'zbek tili uchun bunday tadqiqotlar juda kam, mavjud mashina tarjimasi tizimlari esa o'zbek idiomalarining majoziy ma'nosini to'liq anglab yetmaydi.

Idioma – bu so'zma-so'z tarjima qilinganda ma'nosi chiqmaydigan, faqat majoziy, ko'chma yoki madaniy ma'noda ishlatiladigan barqaror ibora. Ya'ni, idioma – har doim birgalikda ishlatiladigan, o'zining lug'aviy ma'nosidan farqli majoziy ma'no beradigan ibora hisoblanadi. O'zbek tilida frazeologizmlar va maqollar kabi birliklar idiomalarni tashkil qiladi.

Obrazli birliklar – bu tilning o'ziga xos, ko'chma ma'nodagi, tasvir yaratishga xizmat qiluvchi, ma'nosi so'zlarning lug'aviy ma'nosidan kelib chiqmaydigan birikmalardir. Ular nutqqa badiiylilik, ifodaviylilik, hissiy rang kiritadi. Obrazli birliklar odatda metafora, metonimiya va tasviriy idodalar(ramzlar) orqali yuzaga keladi.

Hozirgi kunda mashina tarjimasi takomillashib NMT modellariga o'qitilmoqda. NMT tizimi odatdagi statistik mashina tarjimadan (SMT) farqli ravishda butun gapni kontekst bilan hisobga olgan holda tarjima qiladi. Shu bilan birga, u kam uchraydigan so'zlar, uzun gaplar, domen o'ziga xosligi kabi cheklovlarga ega[1]. NMT – bu tilni bir tildan ikkinchisiga butun jarayonini end-to-end neyron tarmoqlar orqali o'rganadigan tizimdir. Model kiruvchi jumlaning encoder orqali semantik reprezentatsiyaga aylantiradi va decoder orqali maqsad tilida generatsiya qiladi. Shu bilan birga, attention mexanizmi yordamida so'zlar orasidagi moslikni (alignment) o'rganadi[2].

NMT ko'plab yutuqlarga erishgan bo'lsa-da, ularning cheklovlari ham bor – ayniqsa: domen mosligi, trening ma'lumotlari miqdori, kam uchraydigan so'zlar, uzun gaplar, so'zlarni hizalash va beam-search bilan bog'liq muammolar. Bu jihatlar idiomalar va kam-chastotali frazeologik birliklarga yomon ta'sir qiladi[1].

Idiomatik ifodalarni NMT avtomatik tarzda noto'g'ri – ko'pincha literal – tarjima qiladi, shuning uchun idiom xatolarini o'lchash uchun maxsus metrikalar va test-to'plamlar kerak. Mualliflar literal tarjima xatolarini avtomatik aniqlovchi metrikani taklif qiladi va pretraining (monolingual) yordam berishi mumkinligini ko'rsatadi[3].

Idiomalar va obrazli ifodalar mashina tarjimasida eng murakkab leksik-semantik birliklardan biri bo'lib, ular bir nechta chuqur tilshunoslik va kompyuter lingvistika muammolariga sabab bo'ladi.

1. Ko'chma ma'noning aniqlanmasligi. Mashina tarjima tizimlari, ayniqsa NMT modellar, odatda so'zlar ketma-ketligini statistik ehtimollik asosida tarjima qiladi. Idiomalarning esa lug'aviy ma'nosi bilan kontekstual ma'nosi o'rtasida katta farq bor. Kompyuter lingvistik model idiomani "ko'chma birlik" sifatida emas, balki oddiy so'zlar majmui sifatida qabul qiladi. Masalan, google va yandex tarjima tizimlari ushbu iborani quyidagicha tarjima qilgan: Uning ko'zi uchib turibdimi? - Is his eye twitching?

2. Kontekst talab qilishi yoki qilmasligi. Ko'p idiomalar faqat kontekst orqali to'g'ri tushuniladi. NMT tizimlari ba'zan global kontekstni yaxshi ishlay olmaydi, ayniqsa qisqa gaplarda yoki strukturaviy murakkab matnlarda seziladi. Agar tilga ishlov berilganda idiomalar bir maxsus belgiga teglanib ma'no berilgan bo'lsa, ushbu idioma kontekst ichida o'zining

ma'nosini saqlab qolishi yoki aksincha faqat idioma o'zi yolg'iz kelganda ma'nosini saqlashi mumkin. Masalan o'zbek tilidagi idiomalarni 600 dan ortig'ini quyidagicha teglashni amalga oshirdik(1-rasmga qarang).

1-RASM.

3. Madaniy o'ziga xoslik. Idiomalarning katta qismi etno-madaniy, tarixiy, mentalitetga oid birliklardir. Mashina tarjima tizimlari bu madaniy farqlarni chuqur kodlay olmaydi. Masalan, o'zbek tilidagi "yuragi yorildi" – "o'ta xursand bo'ldi" ma'nosida. Inglizchada esa literal tarjima tragediyani ya'ni o'ldi degan ma'noni anglatadi.

4. O'xshash idiomalarning tillararo nomutanosibligi. Bir xil ma'noni anglatuvchi idiomalar ikki tilda ham mavjud bo'lishi lekin ular shaklda turlicha bo'lishi mumkin. Masalan, o'zbekcha: "Og'zi pishgan odam" – tajribali. Inglizcha: "a seasoned person"- tajribali (so'zma-so'z emas).

5. Ko'p variantlilik. Idiomalar ko'pincha bir nechta variantda uchraydi. Masalan, "a'zoyi badandan sovuq ter chiqarib yubormoq" "a'zoyi badandan ter chiqib ketmoq" "badandan ter chiqib ketmoq", "peshanadan sovuq ter chiqmoq", "ter chiqib ketmoq" kabi. NMT modellar har bir variantni yagona semantik blok sifatida emas, balki alohida o'rganadi. Natijada ularning umumiy obrazini aniqlay olmaydi.

6. Past chastotali birliklar. Idiomalarning ko'pi korpuslarda kam uchraydi. NMT modellar esa yuqori chastotali birliklarga tayanadi. Kam uchragan idiomalar semantik jihatdan kamroq o'rganilgan, shuning uchun xatoliklar ko'p bo'ladi.

7. Tasviriy ma'no (imagery) yo'qolib ketishi. Obrazli birliklar tasviriy ifodalar yaratadi. Mashina tarjimasi esa odatda tasviriy ma'noni tan olmaydi, chunki semantik tarmoqlar yetarli boy emas. Natijada tarjimada stilistik, badiiy qiymat yo'qoladi. Biz ushbu muammoga yechim sifatida 337 ta tasviriy ifodalarning kompyuterga o'qitishga tushunarli darajada teglangan lingvistik resurs, baza shakllantirdik. Unda tasviriy ifoda va uning ma'nosi, so'z turkumi, uslubi, nutq turi, bo'yoqdorligi ko'rsatilgan(2-rasmga qarang).

2-RASM.

Demak, ushbu shakllantirilgan leksik resurslarni biz WordNet uchun baza sifatida olishimiz mumkin bo'ladi. WordNet –

tilshunoslik va kompyuter lingvistikasida keng qo'llaniladigan leksik semantik ma'lumotlar bazasi. WordNetdagi so'zlar "synset" (ya'ni "ma'noviy birliklar to'plami")larga guruhlanadi: har bir synsetga bir nechta sinonim so'zlar tegishli, va synsetlar orasida hyperonym–hyponym, antonim, meronym–holonym kabi semantik munosabatlar belgilanadi. Bu struktur – so'zlar va ularning ma'nolari orasidagi semantik tarmoqqa asoslanib, so'z ma'nolarini aniqlash (word sense disambiguation), leksik resurslar yaratish, mashina tarjimasini, tabiiy tilni qayta ishlash (NLP) va boshqa ko'plab vazifalar uchun fundament bo'ladi[4].

O'zbek tilida ushbu sohada tadqiqot o'tkazilgan bo'lib, UzWordNet – WordNet modelining o'zbek tiliga moslangan versiyasidir[5].

UzWordNet 2021 yilda taqdim etilgan bo'lib, unda 28 140 synset, 64 389 sense (so'z ma'nosi varianti), va 20 683 so'z mavjudligini mualliflar bildiradi. Bu o'zbek tilidagi eng keng ko'lamli leksik semantik ma'lumotlar bazasi bo'lib, tilni raqamli resurslarga asoslash, NLP va tarjima tizimlari uchun poydevor yaratishda muhim vosita[5].

WordNet va UzWordNetlarning tuzilishi leksik bog'liqlik (sinonim–antonim), semantik bog'liqlik (hyperonym–hyponym), leksik kategoriyalar (fe'l, ot, sifat ...) asosida tashkil topgan[6].

Shuning uchun, WordNet (va UzWordNet) tilni algoritmik va semantik tahlil qilish, tarjima, so'z ma'nolarini avtomatik aniqlash (WSD), NLP va boshqa kompyuter lingvistika loyihalari uchun nazariy asosi bo'la oladi.

Xulosa qilib aytadigan bo'lsak, o'zbek tilidagi idiomalar, frazeologik birliklar va obrazli ifodalar mashina tarjimasida eng murakkab leksik semantik birliklar hisoblanadi. Ular literal tarjima qilinmaydi, balki kontekst, madaniy moslik va semantik chuqurlikka tayangan holda tarjima qilinadi. NMT modellarining literal tarjimaga moyilligi, kontekstni yetarlicha tushunmasligi va semantik qatlamlarni talqin qilishdagi cheklovlari idiomalarni noto'g'ri tarjima qilishga olib keladi.

Maqolada aniqlangan asosiy muammolar: ko'chma ma'noning aniqlanmasligi, kontekstga bog'liqlik, madaniy xususiyatlar, tillararo nomutanosiblik, ko'p variantlilik, past chastotali birliklar va tasviriy ma'noning yo'qolishi. Ushbu muammolarni hal qilishda WordNet va UzWordNet kabi semantik tarmoqlarni baza sifatida ishlatish va maxsus lingvistik resurslar yaratish samarali yechim bo'lishi mumkin.

Shu bilan birga, maqola o'zbek tilida idiomalarni avtomatik aniqlash va tarjima qilish bo'yicha kompyuter lingvistikasida ilgari surilgan tadqiqotlar uchun nazariy asos yaratadi va amaliy loyihalar uchun yo'l xaritasini belgilaydi.

REFERENCES

1. Koehn P., Knowles R. Six challenges for neural machine translation //arXiv preprint arXiv:1706.03872. – 2017.
2. Bahdanau D. Neural machine translation by jointly learning to align and translate //arXiv preprint arXiv:1409.0473. – 2014.
3. Baziotis C., Mathur P., Hasler E. Automatic evaluation and analysis of idioms in neural machine translation //Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. – 2023. – C. 3682-3700.
4. <https://es.wikipedia.org/wiki/WordNet>
5. Agostini A. et al. Uzwordnet: A lexical-semantic database for the uzbek language //Proceedings of the 11th Global Wordnet conference. – 2021. – C. 8-19.
<https://inlibrary.uz/index.php/crjps/article/view/57587>