

## IBORALARNI AVTOMATIK TARJIMA QILISHDA MODELLASHTIRISHNING AHAMIYATI

**Abjalova Manzura Abdurashetovna**  
Filologiya fanlari doktori, professor  
[abjalova.manzura@gmail.com](mailto:abjalova.manzura@gmail.com)  
ToshDO‘TAU

**Sharipova Sarvinoz,**  
o‘qituvchi  
[shsara232323@gmail.com](mailto:shsara232323@gmail.com)  
SamDCHTI

**Annotatsiya.** Mazkur maqolada zamonaviy tilshunoslik va kompyuter lingvistikasining eng dolzarb muammolaridan biri – iboralarni avtomatik tarjima qilish jarayonida yuzaga keladigan muammolar tizimli tahlil qilinadi. NLP doirasida mashina tarjimasi tizimlarini takomillashtirish hamda parallel korpuslarni yaratish jarayonida iboralarning to‘g‘ri va mos tarjimasini ta‘minlash dolzarb vazifalardan biri sifatida e‘tirof etiladi. Sun‘iy intellekt texnologiyalarining rivojlanishiga qaramay, iboralarning ko‘chma ma‘no qatlamlarini va milliy-madaniy bo‘yoqdorligini mashina tarjimasida to‘g‘ri aks ettirish masalasi hamon murakkabligicha qolmoqda.

**Kalit so‘zlar:** *iboralar tarjimasi, mashina tarjimasi, tabiiy tilni qayta ishlash (NLP), parallel korpus, semantik moslik, lingvokulturologik xususiyatlar.*

**Abstract.** This article provides a systematic analysis of one of the most pressing issues in modern linguistics and computational linguistics – namely, the challenges that arise in the automatic translation of idiomatic expressions. Within the field of Natural Language Processing (NLP), ensuring accurate and contextually appropriate translation of idioms is recognized as a key task, particularly in the development of machine translation systems and the construction of parallel corpora. Despite the rapid advancement of artificial intelligence technologies, effectively capturing the figurative meanings and culturally embedded nuances of idiomatic expressions in machine translation remains a complex and unresolved problem.



**Key words:** *idiom translation, machine translation, natural language processing (NLP), parallel corpus, semantic equivalence, linguocultural features*

Zamonaviy tilshunoslik va tabiiy tilni qayta ishlash (Natural Language Processing – NLP) sohalarida iboralarni, bir tildan ikkinchi tilga avtomatik tarzda tarjima qilish masalasi dolzarb ilmiy muammolardan biri sifatida e'tirof etilmoqda, chunki bunday birliklar o'zining ko'chma ma'noga egaligi, tarkibiy jihatdan bo'linmasligi hamda muayyan lingvokulturologik xususiyatlarni mujassam etishi bilan boshqa leksik birliklardan tubdan farq qiladi. Shu bois, mashina tarjimasi tizimlari tomonidan iboralarni to'g'ri qayta ishlash va ularning semantik ekvivalentlarini aniqlash jarayoni nafaqat texnologik, balki chuqur lingvistik va kognitiv yondashuvni talab etadi.

Ayniqsa, iboralarning leksik tuzilishi bilan ularning haqiqiy semantik mazmuni o'rtasidagi nomutanosiblik avtomatik tarjima tizimlari uchun jiddiy qiyinchilik tug'diradi, chunki aksariyat algoritmlar hali ham ko'p hollarda so'zlar darajasida yoki ehtimollik asosida ishlaydi, bu esa iboralarning noto'g'ri, ya'ni so'zma-so'z tarjima qilinishiga olib keladi. Masalan, ingliz tilidagi *to break a promise* iborasining *va'dani sindirmoq* tarzida tarjima qilinishi uning asl ma'nosi *va'dani buzmoq* bilan mos kelmaydi, bu esa semantik xatolikni yuzaga keltiradi

Frazeologiya va so'z birikmalarining lingvistik tabiati masalalarini chuqur o'rgangan yetuk tadqiqotchilardan biri A.Stefanovich hisoblanadi. Uning ilmiy ishlari frazeologik birliklarni tahlil qilishda korpusga asoslangan yondashuvning muhimligini asoslab berishi bilan alohida ahamiyat kasb etadi. Xususan, u iboralarni faqat leksik birlik sifatida emas, balki yaxlit semantik-struktural sistema sifatida talqin qilib, ularning grammatik va sintaktik xususiyatlarini tizimli ravishda o'rgangan [8:211–213]. Shu bilan birga, A.Stefanovich frazeologik birliklarning turli grammatik kategoriyalar doirasida, jumladan zamon va mayl shakllarida qanday o'zgarishini tahlil qilib, ularning dinamik xususiyatlarini ochib bergan



hamda iboralarni anglash va hosil qilish jarayonida ishtirok etuvchi kognitiv mexanizmlar haqidagi ilmiy qarashlarning rivojlanishiga sezilarli hissa qo'shgan.

Frazeologiya sohada muhim tadqiqotlar olib borgan yana bir nufuzli olim S. Granjer bo'lib, uning korpus lingvistikasi asosidagi ishlari ayniqsa ikkinchi tilni o'rganish jarayonida iboralarning o'zlashtirilishi va qo'llanishini tahlil qilishda keng qo'llanilgan [3:95–98].

Ushbu yondashuv M. Abjalova hamda U. Rashidova tomonidan olib borilgan tadqiqotlarda ham o'z ifodasini topgan bo'lib, ular turk va ingliz tillaridagi frazeologik birliklar asosida parallel korpus yaratishga alohida e'tibor qaratganlar [1:3147]. Mazkur tadqiqotlarda, shuningdek, mashina tarjimasida tizimlarida iboralarning parallel tarjima qilinish jarayonida yuzaga keladigan kamchiliklar atroflicha tahlil qilinib, ularni bartaraf etish bo'yicha ilmiy asoslangan takliflar ilgari surilgan.

Mazkur tadqiqotda tarjima tizimlari orqali matn tarkibida hamda alohida birlik sifatida qo'llangan iboralarning tarjima natijalarini tahlil qilish maqsadida ma'lumotlarni yig'ish (data collection) metodidan foydalanildi. Ushbu jarayonda o'zbek tilidagi iboralar maxsus korpus manbalari asosida jamlanib, ular ustida statistik tahlil amalga oshirildi.

Tadqiqot doirasida Google Translate va Yandex Translate xizmatlari yordamida ingliz tiliga oid bir qator iboralar o'zbek tiliga tarjima qilinib, hosil bo'lgan natijalar lingvistik jihatdan baholandi. Xususan, tarjima jarayonida yuzaga kelgan leksik-grammatik nomuvofiqliklar, semantik xatoliklar hamda stilistik noaniqliklar tizimli ravishda tahlil qilinib, ularning kelib chiqish sabablari aniqlashtirildi.

*1-jadval. Avtomatik tarjima xizmatlarida taqdim etilgan o'zbek-ingliz iboralari tarjimasidagi natijalar.*

№	O'zbek iborasi	Ingliz tilidagi ekvivalenti	Ma'nosi	Tarjimon tizim nomi	Tizimdagi tarjima
1.	Tuyaning dumi yerga tekkanda	When pigs fly	Never	Google	Cho'chqalar uchganda
				Yandex	Cho'chqalar uchganda
2.	Zahardan qimmat turmoq	To cost an arm and a leg	To be very expensive	Google	Qo'l va oyoqqa xarajat qilish
				Yandex	Qo'l va oyoqqa xarajat qilish
3.	Xamirdan qil sug'urgandek	A piece of cake	Very easy	Google	Bir bo'lak kek
				Yandex	Bir parcha pirojnoye
4.	Yaraga tuz sepmaq	Add insult to injury	To make a bad situation worse	Google	Jarohatga haqorat qo'shing
				Yandex	Jarohatga haqorat qo'shing

Mazkur jadvalda o'zbek tilidagi bir qator iboralarning ingliz tilidagi ekvivalentlari, ularning semantik mazmuni hamda avtomatik tarjima tizimlari orqali hosil qilingan natijalar qiyosiy tahlil asosida keltirilgan bo'lib, ushbu ma'lumotlar iboralarni mashina tarjimasida jarayonida yuzaga keladigan muammolarni aniqlash va baholash imkonini beradi. Tahlildan ko'rinadiki, aksariyat hollarda avtomatik tarjima tizimlari iboralarni to'g'ri tushuna olmay, ularni yuzaki strukturaga asoslangan holda, ya'ni so'zma-so'z tarjima qiladi, bu esa semantik va stilistik xatolarning yuzaga kelishiga sabab bo'ladi.

Xususan, “tuyaning dumi yerga tekkanda” iborasi ingliz tilidagi “when pigs fly” iborasi bilan semantik jihatdan mos kelib, “hech qachon” degan ma'noni anglatadi, biroq avtomatik tarjima tizimlari tomonidan “cho'chqalar uchganda” tarzida tarjima qilinishi mazkur birlikning asl semantik mazmunini emas, balki uning yuzaki komponentlarini aks ettirganligini ko'rsatadi. Bu esa tizimlarning



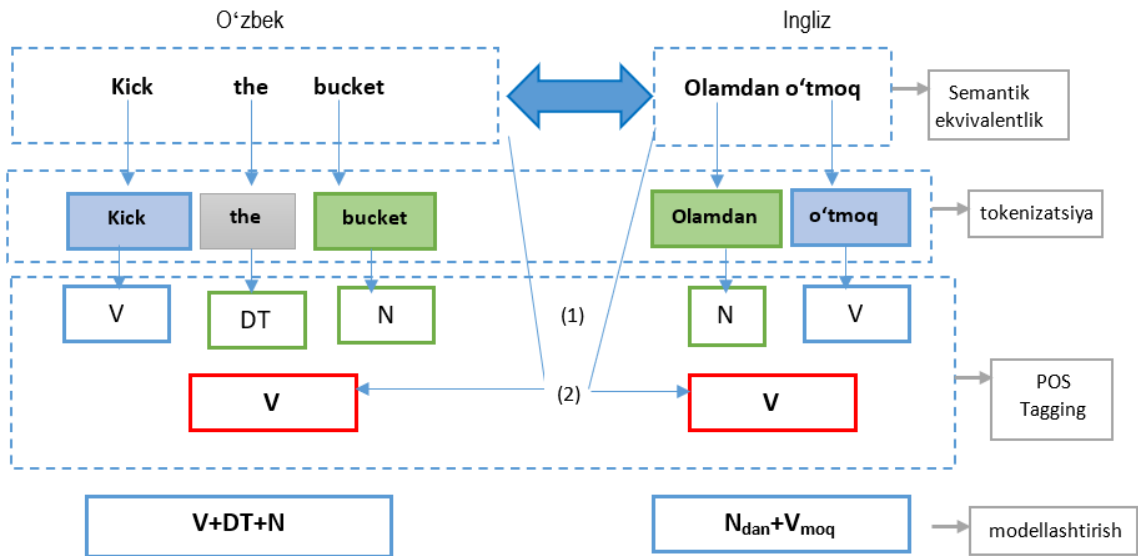
lingvokulturologik moslikni hisobga olishda yetarli darajada samarali emasligini anglatadi.

Ikkinchi “zahardan qimmat turmoq” iborasi ingliz tilidagi “to cost an arm and a leg” iborasiga muqobil bo‘lib, “juda qimmat” ma’nosini bildiradi, biroq avtomatik tarjima natijasida “qo‘l va oyoqqa xarajat qilish” shaklida ifodalanishi iboraning semantik yadrosi noto‘g‘ri talqin qilinganligini ko‘rsatadi.

Keyingi “xamirdan qil sug‘urgandek” iborasi ingliz tilidagi “a piece of cake” ifodasi bilan ma’nodosh bo‘lib, “juda oson” mazmunini anglatadi, biroq avtomatik tarjima natijasida “bir bo‘lak kek” yoki “bir parcha pirojnoye” tarzida berilishi iboraning faqat leksik komponentlari tarjima qilinganini, uning semantik funksiyasi esa butunlay e’tibordan chetda qolganini ko‘rsatadi.

So‘nggi misolda “yaraga tuz sepmoq” iborasi “add insult to injury” konstruktsiyasi bilan mos kelib, “vaziyatni yanada yomonlashtirmoq” ma’nosini bildiradi, ammo avtomatik tarjimada “jarohatga haqorat qo‘shing” kabi shaklda ifodalanishi stilistik va semantik jihatdan noto‘g‘ri bo‘lib, bu holat ham tizimning kontekstual va idiomatik interpretatsiya imkoniyatlari cheklanganligini ko‘rsatadi.

Shundan so‘ng, iboralarning morfologik tarkibini chuqur tahlil qilish hamda avtomatik tarjima jarayonida grammatik va semantik moslikni ta’minlash uchun modellashtirish metodidan foydalanildi.



**1-rasm. O‘zbek-ingliz iboralarini modellashtirish**

Mazkur model transformatsion yondashuv asosida iboralarning strukturaviy-semantik tahlili va tarjima jarayonini bosqichma-bosqich ko‘rsatib beruvchi sxemadir. Unda ingliz tilidagi “kick the bucket” iborasi va uning o‘zbek tilidagi ekvivalenti “olamdan o‘tmoq” o‘rtasidagi semantik moslik izchil ravishda ochib berilgan.

Avvalo, modelda frazeologik birliklar yuzaki struktura darajasida ko‘rib chiqilib, ingliz tilidagi “kick the bucket” iborasi alohida komponentlarga – kick, the, bucket – ajratilgan holda tasvirlangan bo‘lsa, o‘zbek tilidagi mos ekvivalent “olamdan o‘tmoq” iborasi ham strukturaviy jihatdan alohida leksemalarga ajratilgan. Bu bosqich tokenizatsiya jarayonini ifodalaydi, ya’ni murakkab til birliklarini kichik tarkibiy qismlarga ajratish orqali dastlabki tahlil amalga oshiriladi.

Keyingi bosqichda ushbu komponentlarga morfologik va sintaktik teglar (POS tagging) birlashtiriladi, ya’ni ingliz tilida kick (V), the (DT), bucket (N) shaklida belgilansa, o‘zbek tilida olamdan (N) va o‘tmoq (V) tarzida aniqlanadi. Bu jarayon til birliklarining grammatik kategoriyalarini aniqlash orqali ularning funksional rolini belgilashga xizmat qiladi.



Modelning muhim jihati shundaki, u yuzaki strukturadagi ko‘p komponentli birlik (V+DT+N) aslida semantik jihatdan yagona predikativ birlikka tenglashishini ko‘rsatadi; ya’ni ingliz tilidagi uch komponentli konstruktsiya transformatsiya natijasida yagona fe‘l semantikasi (V)ga ega bo‘ladi. Xuddi shuningdek, o‘zbek tilida ham N+V shaklidagi konstruktsiya umumiy holda bitta semantik predikatni ifodalaydi. Bu jarayon modelda (1) va (2) bilan ko‘rsatilgan transformatsion bosqichlar orqali ifodalangan bo‘lib, ular chuqur strukturaga o‘tish va qayta generatsiya qilish jarayonlarini anglatadi.

Shuningdek, modelda “semantik ekvivalentlik” tushunchasi alohida ko‘rsatilib, ikki til o‘rtasidagi moslik aynan yuzaki shaklda emas, balki chuqur semantik darajada ta’minlanishi zarurligi ta’kidlanadi. Ya’ni “kick the bucket” va “olamdan o‘tmoq” iboralari strukturaviy jihatdan farqli bo‘lsa-da, ular umumiy semantik mazmun – “vafot etmoq” – asosida ekvivalent hisoblanadi.

Modelning quyi qismida esa umumlashtirilgan strukturaviy formulalar (V+DT+N va N+V) esa avtomatik tarjima tizimlarida iboralarni aniqlash va ularni to‘g‘ri tarjima qilish uchun muhim metodologik asos bo‘lib xizmat qiladi.

Umuman olganda, mazkur sxema iboralar tarjimasida yuzaki strukturadan chuqur semantik qatlamga o‘tish, so‘ngra esa bir tildan ikkinchi tilga to‘g‘ri shaklda qayta generatsiya qilish zarurligini ilmiy asosda ko‘rsatib beradi hamda avtomatik tarjima tizimlarida tokenizatsiya, morfologik tahlil, POS tagging va semantik modellashtirish kabi bosqichlarning o‘zaro uzviy bog‘liqligini yaqqol namoyon etadi.

### Foydalanilgan adabiyotlar:

1. Abjalova M., Rashidova U., Hasanov A. Creating a phraseological base for the Uzbek↔Turkish automatic translation program. International Scientific and Practical Conference on Actual Problems of Mathematical Modeling and



Information Technology. AIP Conf. Proc. 3147, – pp. 040023-1–040023-5;  
<https://doi.org/10.1063/5.0210705>

2. Abjalova M., E. Adalı and O. Iskandarov, Educational Corpus of the Uzbek Language and its Opportunities, 8th International Conference on Computer Science and Engineering (UBMK), Burdur, Turkiye, 2023, pp. 590-594, doi: 10.1109/UBMK59864.2023.10286682.
3. Granger S., Meunier F. Phraseology: Interdisciplinary Perspective. John Benjamins Publishing, 2008. – 422 p.
4. J. Briskilal, C.N. Subalalitha, An ensemble model for classifying idioms and literal texts using BERT and RoBERTa, Information Processing & Management, 2022.
5. M. Abjalova, U. Rashidova and S. Sharipova, The Importance of Transformational Method in Automatic Idiom Translation, 10th International Conference on Computer Science and Engineering (UBMK), Istanbul, Turkiye, 2025, pp. 1761-1766, doi: 10.1109/UBMK67458.2025.11206800.
6. M. Baker, In Other Words: A Coursebook on Translation (3rd ed.), Routledge, 2018, pp. 65.
7. Naziya Shaikh, Determination of Idiomatic Sentences in Paragraphs Using Statement Classification and Generalization of Grammar Rules, In Proceedings of the WILDRE5– 5th Workshop on Indian Language, 2020.
8. Stefanowitsch A., Stefan Th. G. Collostructions: Investigating the interaction of words and constructions. Interantional Journal of Corpus Linguistics. John Benjamin Publishing Company, 8 (2), 2003. pp. 209-243.
9. Z. Ren, Y. Lu, Q Liu, Y. Huang, Improving statistical machine translation using domain bilingual multiword expressions. In Proceedings of the Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications, 2009, pp. 47–54.



10. Кунин А.В. Курс фразеологии современного английского языка.  
Москва, «Высшая школа» Дубна, Издательский центр «Феникс» 1996.