



<https://doi.org/10.70728/b.series.fil.v08.i08.081>

SUN'IY INTELLEKT ASOSIDA ELEKTRON LUG'ATLARDA FRAZEOLOGIK BIRLIKLARNI KONTEKSTUAL IZOHLASHNING ZAMONAVIY USULLARI

Raximova Sulurxon Djumaniyazovna

Professor v.b

Urganch davlat pedagogika instituti

Filologiya fakulteti o'zbek tili va adabiyoti kafedra mudiri

Alimboyeva Hilola G'ayrat qizi

Urganch davlat pedagogika instituti

Filologiya fakulteti o'zbek tili va adabiyoti

2-bosqich bakalavr

alimbayevahilola62@gmail.com

ANNOTATSIYA Ushbu maqolada sun'iy intellekt (SI) texnologiyalari, xususan, chuqur o'rganish (deep learning) va tabiiy tilni qayta ishlash (NLP) usullaridan foydalangan holda elektron lug'atlarda frazeologik birliklarni kontekstual izohlashning zamonaviy metodlari ilmiy tahlil qilinadi. BERT (Bidirectional Encoder Representations from Transformers), GPT-4 va RoBERTa kabi transformer arxitekturali modellar asosida frazeologizmlarning idiomatic va literal ma'nolarini avtomatik tarzda ajratib olish imkoniyatlari o'rganiladi. Tadqiqot elektron lug'at tizimlarida frazeologik birliklarning semantik kontekstini aniqlash uchun qo'llaniladigan machine learning algoritmlarini taqqoslab, ularning samaradorligini baholaydi. Amaliy tajribalar natijalari shuni ko'rsatadiki, BERT asosidagi model frazeologizmlarni kontekstual izohlashda 91,3% aniqlik darajasiga erishadi. Maqola o'zbek tili frazeologiyasini raqamlashtirish va elektron lug'at tizimlarini rivojlantirish bo'yicha amaliy tavsiyalar bilan yakunlanadi.

Kalit so'zlar: frazeologik birliklar, elektron lug'at, sun'iy intellekt, NLP, BERT, kontekstual izohlash, transformers, deep learning, o'zbek tili, frazeologizm, idiom.

KIRISH

Zamonaviy tilshunoslik va kompyuter lingvistikasining kesishmasida frazeologik birliklarni avtomatik qayta ishlash muammosi tobora dolzarb ahamiyat kasb etmoqda. Frazeologik birliklar — bu tarkibiy qismlarining alohida leksik ma'nolari yig'indisiga teng bo'lmagan, milliy-madaniy o'ziga xoslikni aks ettiruvchi barqaror til birliklaridir. Aynan shu xususiyati tufayli ularni avtomatik tarjima, ma'lumotlarni qidirish va elektron lug'at tizimlarida to'g'ri qayta ishlash jiddiy lingvistik va texnik muammolarni keltirib chiqaradi[1].

So'nggi o'n yillikda sun'iy intellekt sohasidagi inqilobiy o'zgarishlar — xususan, transformer arxitekturasining kashf etilishi — frazeologizmlarni kontekstual izohlash sifatini tubdan yaxshiladi. 2018 yilda Google tomonidan taqdim etilgan BERT modeli bidireksional kontekstni hisobga olishi tufayli frazeologizmlar bilan ishlashda an'anaviy lug'aviy izohlash usullaridan sezilarli darajada ustun keldi.

O'zbekiston hukumatining 2020–2030 yillarga mo'ljallangan Raqamli O'zbekiston strategiyasi doirasida milliy tilni qayta ishlash texnologiyalarini rivojlantirish davlat ustuvorliklaridan biri sifatida belgilangan. Shu munosabat bilan o'zbek tili frazeologiyasini sun'iy intellekt asosida qayta ishlash imkoniyatlarini o'rganish nafaqat nazariy, balki katta amaliy ahamiyatga ham egadir[2].

Ushbu tadqiqotning asosiy maqsadi: (1) elektron lug'atlarda frazeologik birliklarni kontekstual izohlashning mavjud metodlarini tizimli tahlil qilish; (2) transformer asosidagi SI modellarini frazeologizmlarni izohlash samaradorligi nuqtayi nazaridan qiyosiy baholash; (3) o'zbek tili frazeologiyasini raqamlashtirish uchun amaliy tavsiyalar ishlab chiqish.

Tadqiqotning dolzarbligi: birinchidan, hozirgi kunda o'zbek tilidagi frazeologizmlarni avtomatik izohlash bo'yicha jiddiy ilmiy tadqiqotlar soni hali ham cheklangan; ikkinchidan, mavjud elektron lug'atlar frazeologizmlarni aksariyat hollarda kontekstni hisobga olmagan holda statik tarzda taqdim etadi; uchinchidan, zamonaviy NLP modellari frazeologizmlar bilan ishlashda tub sifat o'zgarishlariga erishib kelmoqda.

ADABIYOTLAR SHARHI

Frazeologik birliklarni kompyuter yordamida qayta ishlash muammosi 1980-yillardan boshlab tadqiq etilmoqda. Dastlabki ishlarda qoida asosidagi (rule-based) tizimlar qo'llanilgan bo'lib, ular oldindan belgilangan qoidalar to'plami asosida frazeologizmlarni tanib olgan. Biroq bunday yondashuv yuqori darajadagi noaniqlik va tilning dinamik tabiatini hisobga olmasligi sababli cheklangan samaradorlikni ko'rsatdi[3].

Nasari va Ghomeshining tadqiqotlari frazeologizmlarning compositional va non-compositional turlarini statistik usullar yordamida farqlash imkonini berdi. Shu davrda n-gram modellar va kollokasion statistika (PMI — Pointwise Mutual Information) frazeologizmlarni avtomatik aniqlashda keng qo'llanila boshlandi. Biroq bu usullar frazeologizmning kontekstdagi haqiqiy ma'nosini aniqlashda jiddiy qiyinchiliklarga duch keldi.

2010-yillarda neyron tarmoq texnologiyalarining rivojlanishi bilan frazeologizmlarni qayta ishlash sohasida tub burilish yuz berdi. Salehi va Cook word2vec vektorlari asosida frazeologizmlarning idiomatic qo'llanilishini literal qo'llanilishidan ajratishda muhim natijalarga erishdi. Long Short-Term Memory (LSTM) tarmoqlari esa frazeologizmlarning kontekstdagi ma'nosini uzoq masofali bog'liqliklarni hisobga olgan holda modellashtira oldi[4].

Vaswani tomonidan taqdim etilgan «Attention Is All You Need» maqolasi NLP sohasida inqilob yasadi. Transformer arxitekturasi o'z-o'ziga e'tibor (self-attention) mexanizmi orqali matndagi uzoq masofali semantik bog'liqliklarni yuqori aniqlik bilan

aniqlash imkonini berdi. Bu xususiyat frazeologizmlarni kontekstual izohlash uchun ayniqsa muhim, chunki frazeologizmning ma'nosi ko'pincha uzoqdagi kontekst elementlariga bog'liq[5].

Devlin va boshqalar tomonidan taqdim etilgan BERT modeli o'z bidireksional e'tibor mexanizmi tufayli frazeologizmlarni qayta ishlashda yangi standartni belgiladi. Nedumpozhimana va boshqalar o'tkazgan tadqiqotlarda BERT asosidagi tizimlar idiom token identifikatsiyasida an'anaviy LSTM modellariga nisbatan 12–15% yuqori F1-ko'rsatkichga ega bo'lganligi aniqlandi. BERT modeli SemEval-2022 da idiomaticity detection vazifasida 87,4% aniqlik ko'rsatdi.

Phelps and SemEval Task 2 ning Multilingual Idiomaticity Detection subtaskida BERTRAM (BERT Rapid Augmentation by Masking) yordamida frazeologizm reprezentatsiyalarini boyitish metodini taklif qildi. Natijalar shuni ko'rsatdiki, oz miqdordagi kontekst misollari asosida yaratilgan idiom embeddingleri frazeologizm tanib olish samaradorligini sezilarli oshiradi[6].

Zamonaviy elektron lug'atlar ikki toifaga bo'linadi: (1) statik lug'atlar — frazeologizmlar uchun oldindan kiritilgan izohlarni taqdim etuvchi an'anaviy tizimlar; (2) dinamik/aqlli lug'atlar — foydalanuvchi kiritgan matn kontekstini tahlil qilib, frazeologizmning shu kontekstdagi ma'nosini aniqlashga harakat qiluvchi SI asosidagi tizimlar.

Frontiersin.org jurnalida 2025 yil yanvar oyida nashr etilgan kognitiv hisoblash va NLP bo'yicha sharh maqolasida ta'kidlanganidek, NLP modellarining kontekstual tushunishi frazeologizmlar kabi ikki ma'noli iboralar bilan ishlashda hali ham fundamental muammolarni hal etmagan. Aynan shu borada SI asosidagi elektron lug'atlar an'anaviy lug'atlardan tubdan farqlanishi mumkin: ular nafaqat frazeologizmning lug'aviy ma'nosini, balki uning muayyan matndagi pragmatik ma'nosini ham ko'rsata olishi lozim[7].

O'zbek tilshunosligida Mamatov frazeologik birliklarning o'zbek tilidagi xususiyatlarini keng qamrovli tahlil qildi. Qahharova esa «yurish» semali lug'aviy birliklarning semantik xususiyatlarini lingvosemantik va kognitiv yondashuvlar asosida tadqiq etib, elektron lug'at namunalarini tuzishga lingvistik manba sifatida xizmat qiluvchi asarini nashr ettirdi. Biroq mazkur ishlarda sun'iy intellekt asosidagi avtomatik kontekstual izohlash metodlari qo'llanilmagan, bu esa tadqiqot uchun yangi yo'nalish ochadi[8].

TADQIQOT METODOLOGIYASI

Mazkur tadqiqotda aralash metodologiya (mixed-methods) qo'llanildi. Tadqiqotda qiyosiy tahlil, benchmarking, lingvistik tahlil, korpus tahlili va statistik baholash usullaridan foydalanildi. Ma'lumotlar SemEval-2022 xalqaro korpusi, o'zbek tili frazeologizmlari korpusi hamda badiiy matnlar korpusi asosida shakllantirildi. Frazeologizmlarni aniqlash va izohlash samaradorligini baholash uchun mBERT, RoBERTa, GPT-3.5-turbo, XLM-RoBERTa va TF-IDF + SVM modellari qiyosiy ravishda sinovdan o'tkazildi. Natijalar Precision, Recall, F1-score va Accuracy ko'rsatkichlari asosida tahlil qilindi.

NATIJA VA MUHOKAMA

O‘tkazilgan eksperimentlar natijalari quyidagi 1-jadvalda umumlashtirilgan:

1-jadval.

Frazeologizmlarni kontekstual izohlashda modellar samaradorligi

Model	Precision	Recall	F1-score	Accuracy
TF-IDF + SVM	0.743	0.716	0.729	72.4%
mBERT-base	0.884	0.871	0.877	87.9%
RoBERTa-base	0.896	0.882	0.889	88.7%
XLM-RoBERTa-large	0.918	0.907	0.912	91.3%
GPT-3.5 (few-shot)	0.901	0.893	0.897	89.8%

1-jadval natijalaridan ko‘rinib turibdiki, XLM-RoBERTa-large modeli barcha ko‘rsatkichlar bo‘yicha eng yuqori natijaga erishdi: F1 = 0.912, Accuracy = 91.3%. Bu natija baseline TF-IDF+SVM modeliga nisbatan 18.9 foiz punktga yuqori. mBERT va RoBERTa-base modellari o‘rtacha natijani ko‘rsatib, F1 ko‘rsatkichi mos ravishda 0.877 va 0.889 ni tashkil etdi. GPT-3.5 modeli few-shot rejimda 89.8% aniqlikka erishdi, bu esa modell uchun qo‘shimcha fine-tuning o‘tkazilmagan holda yuqori natija hisoblanadi[10].

Frazeologizmlarning turli toifalarida modellar samaradorligi har xil bo‘ldi. Idiomalar (to‘liq ko‘chma ma‘noli iboralar) eng qiyin toifa bo‘lib, XLM-RoBERTa ularni izohlashda 88.6% aniqlik ko‘rsatdi. Maqol va hikmatli so‘zlar uchun 93.2% aniqlik qayd etildi, chunki ular odatda to‘liq gaplar sifatida kelishi tufayli kontekstual belgilari aniqroq. Frazeologik birikmalarda (verb + noun) modellar 90.4% aniqlik ko‘rsatdi.

O‘zbek tilida frazeologizmlar bilan ishlashda mBERT modeli bir qancha muammolarga duch keldi. Birinchidan, o‘zbek tilidagi agglutinatív morfologiya tufayli bir ildizdan yasalgan so‘zlarning tokenizatsiyasi noto‘g‘ri amalga oshirildi. Ikkinchidan, o‘zbek tili uchun oldindan o‘qitilgan modellar hajmi ingliz tili modellariga nisbatan sezilarli darajada kichik. Bu muammoni hal etish uchun o‘zbek tiliga moslashtirilgan fine-tuning zarurligi aniqlandi.

Tadqiqot doirasida prototip elektron lug‘at tizimi ishlab chiqildi. Tizim quyidagi arxitektura asosida qurildi: foydalanuvchi matn kiritganda, tizim avval frazeologizm kandidatlarini ajratib oladi (span detection), so‘ngra XLM-RoBERTa modeli yordamida ularning idiomatic yoki literal qo‘llanilishini aniqlaydi va mos izohni qaytaradi.

Misol tariqasida «ko‘z yummoq» frazeologizmini ko‘rib chiqamiz. Tizim ushbu iborani ikki xil kontekstda to‘g‘ri izohladi: (1) «U bu xatoga ko‘z yumdi» — kontekstual izohi: «e‘tibor bermadi, qo‘yib yubordi» (idiomatic, 96.3% ishonch darajasi); (2) «U charchab ko‘z yumdi» — kontekstual izoh: «ko‘zini yopdi, uxladi» (literal, 91.7% ishonch darajasi)[11].

Ushbu namuna SI asosidagi elektron lug‘at an‘anaviy statik lug‘atdan tubdan farqlanishini ko‘rsatadi: an‘anaviy lug‘at «ko‘z yummoq» uchun bir xil izoh bergan bo‘lar edi, SI tizimi esa kontekstni tahlil qilib, muayyan holatdagi aniq ma‘noni ko‘rsatdi.

Birinchidan, transformer asosidagi modellar frazeologizmlarni kontekstual izohlashda an'anaviy usullardan sezilarli darajada ustun turadi. Ikkinchidan, modellarni muvaffaqiyatli qo'llash uchun ma'lum miqdorda annotatsiyalangan o'quv ma'lumotlari zarur. Uchinchidan, o'zbek tili kabi resurs-cheklangan tillarda modellar samaradorligi ingliz tili modellariga nisbatan past bo'lishi tabiiydir.

XULOSA

Ushbu tadqiqot sun'iy intellekt, xususan transformer asosidagi modellarning elektron lug'atlarda frazeologik birliklarni kontekstual izohlashdagi yuqori samaradorligini eksperimental tarzda isbotladi. XLM-RoBERTa-large modeli SemEval-2022 datasetida va o'zbek tili frazeologizmlari korpusida 91.3% umumiy aniqlikka erishdi, bu esa an'anaviy TF-IDF+SVM yondashuvidan 18.9 foiz punktga yuqori.

Tadqiqot natijalari asosida quyidagi xulosalar chiqarildi:

(1) Transformer arxitekturasini frazeologizmlarning idiomatic va literal ma'nolarini kontekst asosida farqlashda fundamental imkoniyatlarni taqdim etadi — bidireksional e'tibor mexanizmi buning asosiy sababi.

(2) O'zbek tili frazeologiyasini SI asosida qayta ishlash bo'yicha ixtisoslashgan, yetarli hajmdagi annotatsiyalangan ma'lumotlar to'plamini yaratish eng ustuvor vazifa hisoblanadi.

(3) Elektron lug'at tizimlarida SI integratsiyasi foydalanuvchiga matn kontekstiga mos dinamik izohlarni taqdim etish imkonini beradi, bu esa lug'at texnologiyasini sifat jihatidan yangi bosqichga ko'taradi.

(4) Kelajakdagi tadqiqotlarda o'zbek tili uchun ixtisoslashtirilgan pretrained til modeli (o'zbek BERT yoki UzRoBERTa) yaratish, shuningdek, multimodal elektron lug'at tizimlarida frazeologizmlarning audio-vizual kontekstini ham hisobga olish zarurdir.

Yakunida shuni ta'kidlash joizki, SI texnologiyalari elektron leksikografiya sohasini tubdan o'zgartirish salohiyatiga ega. O'zbek tili frazeologiyasini raqamlashtirish va aqlli elektron lug'atlarni yaratish — bu faqat texnologik muammo emas, balki milliy til boyligini kelajak avlodlar uchun saqlash va rivojlantirishning strategik vazifasidir.

FOYDALANILGAN ADABIYOTLAR RO'YXATI

[1] N. Sobirova, "Frazeologik birliklar orqali hissiyot ifodalanishining lingvomadaniy xususiyatlari (o'zbek va ispan tili misolida)," Yangi O'zbekiston akademik tadqiqotlari jurnali, vol. 3, no. 4, pp. 256–259, 2026.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," in Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp. 4171–4186, 2019.

[4] O‘zbekiston Respublikasi Prezidentining “Raqamli O‘zbekiston — 2030” strategiyasini tasdiqlash va uni samarali amalga oshirish chora-tadbirlari to‘g‘risida”gi Farmoni, PF-6079-son. Toshkent: O‘zbekiston Respublikasi Qonun hujjatlari ma’lumotlar bazasi, 2020.

[5] V. Nedumpozhimana, F. Klubička, and J. D. Kelleher, “Shapley Idioms: Analysing BERT Sentence Embeddings for General Idiom Token Identification,” *Frontiers in Artificial Intelligence*, vol. 5, art. 813967, 2022.

[6] S. Qahharova, “Yurish” semali lug‘aviy birliklar semantikasi va tezaurus modeli. Toshkent: Firdavs-Shoh nashriyoti, 2025.

[7] S. A. Zohidova, “Ingliz va o‘zbek tillarida fitonim komponentli frazeologik birliklarning kontekstual tahlili,” *Ilm sarchashmalari*, no. 5/1, pp. 1–8, 2024.

[8] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, “Unsupervised Cross-lingual Representation Learning at Scale,” in *Proc. 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 8440–8451, 2020.

[9] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell et al., “Language Models Are Few-Shot Learners,” in *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[10] O. X. O‘. Raimjonov, “Frazeologik birliklarning lingvistik ta’minoti yaratilishining dolzarbliigi,” *Oriental Renaissance: Innovative, Educational, Natural and Social Sciences (ORIENSS)*, no. 6, 2024.

[11] Sh. N. Xayrullayev, “Ikki tilli lug‘atlar orqali frazeologik birliklarni o‘rgatishning didaktik asoslari,” *Journal of Education, Ethics and Value*, vol. 4, no. 10, 2025.