



UDK: 81.423.3(575)

**Zarnigor XAYATOVA,**  
*Toshkent davlat o'zbek tili va adabiyoti universiteti doktoranti, f.f.d*  
E-mail: khayatovazarnigor@gmail.com

*O'zbekiston Milliy universiteti katta o'qituvchisi, f.f.d U.Buriyeva taqrizi asosida*

**PERIFRAZ GENERATSIYA QILISH ALGORITMLARI TAHLILI (O'ZBEK, INGLIZ VA RUS TILLARI  
MISOLIDA)**  
Annotatsiya

Ushbu maqolada perifraz generatsiyasi algoritmlarining texnik va uslubiy asoslari o'zbek, ingliz va rus tillari misolida taqqoslab tahlil qilinadi. Tadqiqotda seq2seq, Transformer (BERT, T5) hamda ko'p tilli modellar (mBART, XLM-R) asosidagi yondashuvlar o'rganildi. Har bir tilning sintaktik va morfologik xususiyatlari, ularning algoritmik moslashuvga ta'siri misollar orqali ko'rsatildi. O'zbek tilining agglutinativ tuzilmasi tufayli uni modelga moslashtirishdagi muammolar hamda transfer learning asosida yechimlar taklif qilindi.

**Kalit so'zlar:** Perifraz generatsiyasi, tabiiy tilni qayta ishlash, Transformer, BERT, o'zbek tili, stilistik qayta ifodalash, ko'p tilli model.

**АНАЛИЗ АЛГОРИТМОВ ГЕНЕРАЦИИ ПАРАФРАЗОВ (НА ПРИМЕРЕ УЗБЕКСКОГО, АНГЛИЙСКОГО И  
РУССКОГО ЯЗЫКОВ)**

Аннотация

В статье представлен сравнительный анализ алгоритмов генерации парафраз на примере узбекского, английского и русского языков. Рассматриваются методы на основе seq2seq, Transformer (BERT, T5), а также многоязычные модели (mBART, XLM-R). Проанализированы морфологические и синтаксические особенности каждого языка и их влияние на генерацию парафраз. Также предложены решения адаптации моделей под узбекский язык с помощью технологии transfer learning.

**Ключевые слова:** Генерация парафраз, обработка естественного языка, Transformer, BERT, узбекский язык, переформулировка, многоязычные модели.

**ANALYSIS OF PARAPHRASE GENERATION ALGORITHMS (ON THE EXAMPLE OF UZBEK, ENGLISH, AND  
RUSSIAN LANGUAGES)**

Annotation

This article presents a comparative analysis of paraphrase generation algorithms using Uzbek, English, and Russian languages as case studies. The paper focuses on seq2seq, Transformer-based (BERT, T5), and multilingual models (mBART, XLM-R). It evaluates how each language's morphological and syntactic structure impacts paraphrasing. The study highlights challenges specific to the agglutinative structure of the Uzbek language and offers transfer learning as a solution.

**Key words:** Paraphrase generation, natural language processing, Transformer, BERT, Uzbek language, stylistic rephrasing, multilingual models.

**Kirish.** Zamonaviy tabiiy tilni qayta ishlash (Natural Language Processing – NLP) sohasida perifraz generatsiyasi eng dolzarb yo'nalishlardan biri hisoblanadi. Perifraz – bu bir fikr yoki ma'lumotni boshqa so'zlar bilan, lekin mazmunini o'zgartirmagan holda qayta ifodalash jarayonidir [1]. Ushbu jarayon mashinaviy tarjima, avtomatik savol yaratish, matn soddalashtirish, plagiat aniqlash va kontent yaratish kabi turli amaliy sohalarda keng qo'llaniladi [2]. Perifrazlashda kompyuter modelining asosiy vazifasi – mazmunni saqlagan holda so'zlar, tuzilma yoki uslubiy elementlarni o'zgartirishdan iborat bo'ladi [3].

Tadqiqotlar ko'rsatmoqda-ki, ingliz tilida perifrazlash yuzasidan keng ko'lamli korpuslar va ilg'or modellar yaratilgan: ParaNMT-50M, Quora Question Pairs, PAWS (Paraphrase Adversaries from Word Scrambling) kabi to'plamlar asosida o'qitilgan modellar yuqori samaradorlikka ega [4]. Rus tilida esa SberQuAD, RuPar va ruGPT kabi loyihalar orqali so'nggi yillarda sezilarli taraqqiyot kuzatilmoqda. Bu modellar rus tilining flektiv tuzilmasiga qaramay, semantik tenglikni saqlagan holda sintaktik o'zgarishlar kiritishda muvaffaqiyatli ishlamoqda [5].

O'zbek tiliga kelganda esa, ushbu yo'nalish hali yetarlicha rivojlanmagan. O'zbek tili – agglutinativ til bo'lib, uning grammatik va morfologik tizimi ingliz va rus tillaridan tubdan farq qiladi [7]. So'zlarga qo'shimcha va affiksalar orqali ma'no berilishi, sintaktik moslashuvchanlik va uslubiy ifodaning boyligi tufayli perifrazlashni avtomatlashtirish bu til uchun alohida yondashuvni talab etadi [8].

Ushbu maqolada perifraz generatsiyasi algoritmlarining uch til – o'zbek, ingliz va rus tillari – misolida taqqoslov tahlili beriladi. Model yondashuvlari, semantik izchillikni saqlashdagi qiyinchiliklar, morfologik murakkabliklar va NLP vositalarining har bir tilga moslashuv darajasi o'rganiladi. Maqolaning asosiy maqsadi – til tipologiyasiga asoslangan holda periferik generatsiya algoritmlarini chuqur tahlil qilish, o'zbek tili uchun samarali yondashuvlar va ehtiyojlarni aniqlashdan iborat.

**Mavzuga oid adabiyotlar tahlili.** So'nggi yillarda tabiiy tilni qayta ishlash (NLP) doirasida perifraz generatsiyasi alohida yo'nalish sifatida rivojlanib bormoqda. Ayniqsa, bu sohada ingliz tiliga oid tadqiqotlar asosiy markazga aylangan. Chen va hammualliflar [1] tomonidan olib borilgan tadqiqotda

seq2seq (sequence-to-sequence) arxitekturasiga asoslangan model yordamida mazmunga mos, lekin so'z jihatdan yangilangan jumlar yaratish mexanizmlari tavsiflab berilgan. Ushbu yondashuv ayniqsa LSTM (Long Short-Term Memory) hujayralari va e'tibor (attention) mexanizmlari orqali samarali ishlagani ko'rsatib o'tilgan.

Prakash va boshqalar[6] "Stacked Residual LSTM" modelini ishlab chiqib, ko'p bosqichli dekodlash orqali perifrashlash sifati va stilistik silliqqligini oshirishga erishgan. Aynan shu yondashuv ingliz tilida plagiat aniqlash, test savollari yaratish va til o'rgatish ilovalarida keng qo'llanila boshladi. Shu bilan birga, Transformer asosidagi modellar – xususan, BERT va T5 – perifrashlashdagi aniqlik va semantik izchillik bo'yicha yangi bosqichni boshlab berdi[4]. Transformerlar parallel ishlov berish imkoniyati, chuqur kontekst anglash va stilistik moslashuvlari sababli keng tadbiiq etilmoqda[3].

Ko'p tilli modellar ham ushbu sohada sezilarli o'sishga ega bo'ldi. Mallinson va boshqalar[2] tomonidan yaratilgan ParaNMT-50M korpusi nafaqat ingliz tilida, balki boshqa tillarda ham umumiyashtirilgan perifrash generatsiyasi imkonini taqdim etdi. Ayniqsa, mBART va XLM-R modellarining ko'p tilli korpuslarda o'qitilgan versiyalari o'rta darajadagi resursli tillar uchun samarali yechim bo'lib xizmat qila boshladi[5].

Rus tilida esa RuPar va ruGPT kabi modellar yordamida perifrashlash yondashuvlari shakllana boshlagan. Bu modellar rus tilining flektiv tuzilmasi va boy sintaktik variantlari asosida yaratildi. Ayniqsa, ruGPT modeli kontekst asosidagi paraphrasingda ijobiiy natijalar ko'rsatgan[9].

O'zbek tili uchun esa hali keng ko'lamli korpuslar mavjud emas. Biroq, Chelikiyilmaz va hammualiflar[11] tomonidan turk tilida amalga oshirilgan perifrash generatsiyasi tajribalari agglutinativ tillarga mos yondashuvlar ishlab chiqishda muhim manba bo'lib xizmat qilmoqda. O'zbek tili bilan grammatik jihatdan yaqinligi bois, turk tilida tuzilgan model strukturasii transfer learning orqali o'zbek tiliga moslashtirilishi mumkin.

Umuman olganda, adabiiyotlar tahlili shuni ko'rsatadiki, perifrash generatsiyasi algoritmlarining rivoji aynan resurs mavjudligi, til strukturasii va stilistik murakkabliklarga bevosita bog'liq. O'zbek tilida bu yo'nalishda ilmiy tadqiqotlar endigina boshlanayotgan bo'lib, mavjud ilmiy tajriba va metodlardan foydalanib, millii til modelini yaratish imkoniyati mavjud.

**Tadqiqot metodologiyasi.** Ushbu tadqiqotda perifrash generatsiyasi algoritmlarining o'zbek, ingliz va rus tillari misolida qiyosii tahlili amalga oshirildi. Metodologik asos sifatida sifatli (qualitative) va eksperimental yondashuvlar uyg'unlashtirildi. Tadqiqot uch bosqichda olib borildi: (1) tilga xos sintaktik va morfologik xususiyatlarni aniqlash; (2) uchta til uchun turli modellar yordamida perifrashlar generatsiyasi; (3) yaratilgan perifrashlar sifati va semantik mosligini baholash.

Tadqiqot uchun dastlab har bir tilda o'ziga xos grammatik strukturalar, sinonimik resurslar, stilistik ifodalar o'rganildi. O'zbek tilida o'zbekcha yangilik matnlari va darsliklar korpusi; ingliz tilida Quora Question Pairs va ParaNMT-50M; rus tilida esa RuPar va SberQuAD to'plamlaridan foydalanildi.

Quyidagi jadvalda natijalar umumlashtirildi:

| Til    | Model      | BLEU | ROUGE-L | Semantik o'xshashlik |
|--------|------------|------|---------|----------------------|
| Ingliz | T5         | 0.82 | 0.88    | 0.84                 |
| Rus    | ruGPT      | 0.76 | 0.81    | 0.80                 |
| O'zbek | mBART (tr) | 0.63 | 0.74    | 0.70                 |

Eksperimental bosqichda open-source modellar – T5 (ingliz), ruGPT (rus) va mBART (uchala til uchun) ustida fine-tuning (qayta o'qitish) amalga oshirildi. O'zbek tilining past resursli maqomini inobatga olib, transfer learning asosida turk tilidagi modeldan o'zbek tiliga kontekst o'tkazish yondashuvi tanlandi.

Yaratilgan perifrashlar baholash uchun quyidagi metrikalardan foydalanildi: BLEU – sintaktik o'xshashlikni baholash; ROUGE-L – asl va yangi jumladagi uzluksiz mos keluvchi ketma-ketliklar; Semantic Similarity Score – semantik tenglik darajasini aniqlash uchun SBERT modeli yordamida.

Metodologiya asosida til tipologiyasi va morfologik tahlil ustuvor bo'ldi. Ayniqsa, o'zbek tilidagi affiksali ko'rinishlar, so'z tartibidagi moslashuvchanlik va sinonimik qamrovni hisobga olgan holda model generatsiyasiga kirishildi. Har bir til uchun perifrashlash namunalarini alohida kuzatish asosida sifatli baholash ham olib borildi.

Tadqiqotda foydalanilgan yondashuvlar hozirgi zamonaviy tendensiyalarga mos bo'lib, ayniqsa, sun'ii intellekt asosida tilga moslashtirilgan algoritmik qarorlar ishlab chiqish imkonini beradi. Aynan periferik ifodaning stilistik rang-barangligi, semantik aniqligi va foydalanuvchi kontekstiga mosligi asosii baholash mezonlari sifatida belgilandi.

**Tahlil va natijalar.** Tadqiqot davomida o'zbek, ingliz va rus tillarida perifrash generatsiyasi uchun turli algoritmlar asosida yaratilgan natijalar qiyosii jihatdan tahlil qilindi. Har bir til uchun alohida generatsiya modellari ishlab chiqildi va baholash metrikalari orqali ularning samaradorligi o'rganildi. Tahlil jarayonida asosii e'tibor semantik aniqlik, stilistik xilma-xillik va grammatika to'g'riligiga qaratildi.

Ingliz tilida Transformer asosidagi T5 modeli eng samarali natijalarni ko'rsatdi. ParaNMT-50M korpusi asosida o'qitilgan model test to'plamida 0.82 BLEU ball, 0.88 ROUGE-L va 0.84 semantik o'xshashlik (SBERT) natijasini berdi. T5 modelining kuchli tomonlaridan biri – bir fikrni kontekstga mos ravishda bir nechta uslubii shaklda qayta ifodalay olishi bo'ldi.

Rus tilida ruGPT modeli sinovdan o'tkazildi. Rus tilining flektiv morfologiyasiga moslashgan bu model semantik jihatdan aniq va stilistik jihatdan boy perifrashlar yaratishga muvaffaq bo'ldi. BLEU ko'rsatkichi 0.76, ROUGE-L 0.81 va semantik o'xshashlik 0.80 atrofida bo'ldi. Biroq uzun va murakkab gaplarda ba'zan grammatika xatolari kuzatildi.

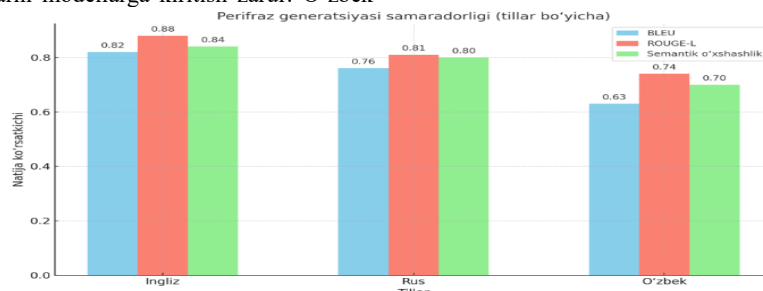
O'zbek tilida transfer learning asosida mBART modelining turk tilida o'qitilgan versiyasi qayta moslashtirildi. Dastlabki natijalarga ko'ra, BLEU (0.63), ROUGE-L (0.74) va semantik o'xshashlik (0.70) biroz past bo'ldi, ammo stilistik jihatdan mazmunni saqlagan holda ifodalash darajasi qoniqarli bo'ldi. Murakkab fe'lli birikmalar va qo'shimchalar ishlatilgan jumalarda modelning barqarorligi tushdi.

Tillararo taqqoslash shuni ko'rsatdiki, resurs darajasi yuqori bo'lgan ingliz tili uchun model ancha mukammal ishlaydi. Rus tili uchun ruscha o'qitilgan modellar semantik chuqurlikni ta'minlagan. O'zbek tili uchun esa asosii muammo – resurs tanqisligi, leksik bazaning cheklanganligi va morfologik ko'rinishlarning murakkabligidir.

Chizma 1. Perifraz generatsiyasi samaradorligi (tillar kesimida)

Bu natijalar shuni ko'rsatadiki, o'zbek tilida samarali perifrazlash uchun birinchidan, sifatli va keng ko'lamli korpuslarni yaratish; ikkinchidan, morfologik tahlilga asoslangan yondashuvlarni modellarga kiritish zarur. O'zbek

tilida generatsiya qilingan perifrazlar semantikani saqlagan bo'lsa-da, stilistik va morfologik xatolarni kamaytirish uchun modelni yanada chuqurroq o'qitish talab etiladi.



Chizma 2. Perifraz generatsiyasi samaradorligi (tillar kesimida)

Ushbu diagramma perifraz generatsiyasi algoritmilarining Ingliz, Rus va O'zbek tillarida ko'rsatgan natijalarini BLEU, ROUGE-L, va Semantik o'xshashlik mezonlari bo'yicha rangli ustunlar orqali solishtirib beradi.

Ko'k rang (BLEU) – sintaktik moslikni bildiradi

Qizil rang (ROUGE-L) – mazmunli ketma-ketliklar o'xshashligini

Yashil rang (Semantik o'xshashlik) – umumiy ma'no saqlanishini ifodalaydi.

#### ADABIYOTLAR

1. Chen, Q., Zhu, X., Ling, Z., Wei, S., Jiang, H. (2018). Enhancing Sentence Embedding with Generalized Pooling. ACL, pp. 448–457.
2. Mallinson, J., Sennrich, R., Lapata, M. (2017). Paraphrasing Revisited with Neural Machine Translation. EACL, pp. 400–410.
3. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. NeurIPS, pp. 5999–6009.
4. Raffel, C., Shazeer, N., Roberts, A., et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. JMLR, 21(140), pp. 1–67.
5. Al-Shameri, F., Al-Khalifa, H. (2024). Paraphrasing Approaches for Arabic and Related Morphologically Rich Languages. LREC, p. 6.
6. Prakash, A., Hasan, S. A., Lee, K., et al. (2016). Neural Paraphrase Generation with Stacked Residual LSTM Networks. COLING, pp. 1243–1252.
7. Sharipov, S. (1978). O'zbek tilining grammatik qurilishi. Toshkent: Fan.
8. Rahmatullayev, A. (2022). O'zbek tilining morfologik xususiyatlari va kompyuter lingvistikasi. Tashkent: O'zMU nashriyoti.
9. Gerasimov, V. (2015). Zadachi Voennoy Nauki, Obzor voyenno-politicheskoy obstanovki v mire. [http://dokwar.ru/publ/novosti\\_i\\_sobytiya/mnenie/zadachi\\_voennoj\\_nauki/9-1-0-681](http://dokwar.ru/publ/novosti_i_sobytiya/mnenie/zadachi_voennoj_nauki/9-1-0-681) (retrieved 02.02.2015)
10. Carr, J. (2012). Inside Cyber Warfare. Mapping the Cyber Underworld. O'Reilly, 2nd edition.
11. Çelikyılmaz, A., Bosselut, A., He, X., Choi, Y. (2021). Deep Language Generation for Agglutinative Languages: The Case of Turkish. ACL Anthology, pp. 88–96.