

UDK: 811.512.133`42

PERIFRAZ GENERATSIYA QILISH ALGORITMLARI TAHLILI

Xayatova Zarnigor Marufovna

doktorant DSc

khayatovazarnigor@gmail.com

ToshDO`TAU

Annotatsiya. Ushbu maqolada perifraz generatsiyasi algoritmlarining nazariy asoslari va texnik jihatlarini o'rganiladi. Tadqiqot davomida seq2seq, Transformer va ko'p tilli modellar (masalan, BERT, T5, mBART) asosidagi yondashuvlar taqqoslanadi. O'zbek tilining morfologik va sintaktik xususiyatlari tahlil qilinib, o'zbek tilida perifraz yaratishda yuzaga keladigan muammolar va istiqbolli yondashuvlar muhokama qilinadi. Real misollar asosida o'zbek tilida perifrazlashning samarali modellari tavsiya etiladi. Maqola natijalari tilshunoslik, NLP va ta'lim sohalarida qo'llanilishi mumkin bo'lgan amaliy yechimlarni taklif qiladi.

Abstract. This article explores the theoretical foundations and technical aspects of paraphrase generation algorithms. The study compares approaches based on seq2seq, Transformer, and multilingual models such as BERT, T5, and mBART. Special focus is given to the morphological and syntactic characteristics of the Uzbek language, along with challenges and promising methods for generating paraphrases in Uzbek. Effective models are proposed through real examples. The results of the article offer practical solutions applicable in linguistics, natural language processing (NLP), and education.

Аннотация. В данной статье рассматриваются теоретические основы и технические аспекты алгоритмов генерации парафраз. В ходе исследования проводится сравнительный анализ подходов, основанных на seq2seq, Transformer и многоязычных моделях (BERT, T5, mBART). Особое внимание уделяется морфологическим и синтаксическим особенностям узбекского языка, а также трудностям и перспективам генерации парафраз на узбекском языке. На основе реальных примеров предлагаются эффективные модели перефразирования. Результаты статьи могут быть полезны в прикладных задачах лингвистики, обработки естественного языка (NLP) и образования.

Kalit so'zlar: *perifraz generatsiyasi, tabiiy tilni qayta ishlash, seq2seq model, Transformer, BERT, T5, mBART, ko'p tilli model, o'zbek tili, semantik izchillik.*

Kirish. Perifraz (ya'ni, bir fikrni boshqa so'zlar bilan qayta ifodalash) tabiiy tilni qayta ishlash (Natural Language Processing – NLP) sohasida muhim

muammolardan biri hisoblanadi. Ushbu sohada tadqiqotlar ko'p jihatdan mashinaviy tarjima, matnni soddalashtirish, plagiat aniqlash va avtomatik savol yaratish kabi qo'llanmalarga yo'naltirilgan. Avvalambor, ingliz tilidagi perifrasiyalar generatsiyasi algoritmlari bo'yicha keng ko'lamli tadqiqotlar olib borilgan. Masalan, tadqiqotlarda chuqur o'rganilgan seq2seq (sequence-to-sequence) modellari yordamida perifrasiyalar yaratish mexanizmlari tahlil qilingan [6]. Tadqiqotda LSTM (Long Short-Term Memory) tarmoqlari va e'tibor (attention) mexanizmlarining samaradorligi ko'rsatib o'tilgan.

Shuningdek, taklif qilingan “Stacked Residual LSTM” arxitekturasida perifrasiyalar sifatini oshirishda muhim rol o'ynagan [7]. Tadqiqotchilar tarafidan ishlab chiqilgan model ko'p bosqichli o'qitish orqali yanada aniq va kontekstga mos jummalarni generatsiya qilishga erishgan. So'nggi yillarda Transformer asosidagi yondashuvlar – xususan, BERT (Bidirectional Encoder Representations from Transformers) va T5 (Text-to-Text Transfer Transformer) modellari perifrasiylashda inqilobiy natijalarni ko'rsatdi [7]. Bu modellarning kuchli tomonlaridan biri – ular katta hajmdagi korpuslar ustida o'qitilib, kontekstni chuqur anglay oladi.

Turli tillarda, xususan, fransuz, nemis va ispan tillarida ham perifrasiylash bo'yicha tadqiqotlar mavjud. Masalan, Mallinsonlar [10] o'z tadqiqotida ko'p tilli perifrasiyalar korpuslari yordamida neyron tarmoqlarning umumiyashtirish (generalization) qobiliyatini oshirishga erishgan. Ular tomonidan taqdim etilgan “ParaNMT-50M” korpusi ko'plab neyron modelni qayta o'qitish va sinovdan o'tkazishda ishlatilmoqda. Afsuski, o'zbek tilida perifrasiyalar generatsiyasi bo'yicha tadqiqotlar hozircha cheklangan. Biroq turkiy tillar, xususan, turk tilidagi yondashuvlar ma'lum darajada o'zbek tili uchun andoza bo'lishi mumkin. Çelikiyilmaz [11] tomonidan turk tilida ishlab chiqilgan periferik matnlar generatsiyasi tizimi, ayniqsa, semantik izchillikni saqlashda ancha samarali ekani qayd etilgan. Shunday qilib, adabiyotlar sharhi shuni ko'rsatadiki, perifrasiyalar generatsiyasi bo'yicha algoritmlar izchil rivojlanib bormoqda. Transformer arxitekturasida bu borada yangi bosqichni boshlab berdi. Biroq o'zbek tilida bu sohada hali yetarlicha tadqiqot olib borilmagan bo'lib, bu yo'nalishda izlanishlar boshlanishi dolzarb va istiqbolli hisoblanadi.

Algoritmlarning texnik tahlili

Perifrasiyalar generatsiyasi bo'yicha eng ko'p qo'llaniladigan yondashuvlar uch asosiy guruhga ajraladi: birinchisi - an'anaviy neyron tarmoqlar, ikkinchisi - Transformer asosidagi zamonaviy arxitekturalar, va uchinchi - ko'p tilli modellar. Har bir guruh o'ziga xos texnik xususiyatlarga, ishlash mexanizmlariga va vazifaga bog'liq ustunlik yoki cheklavlarga ega.

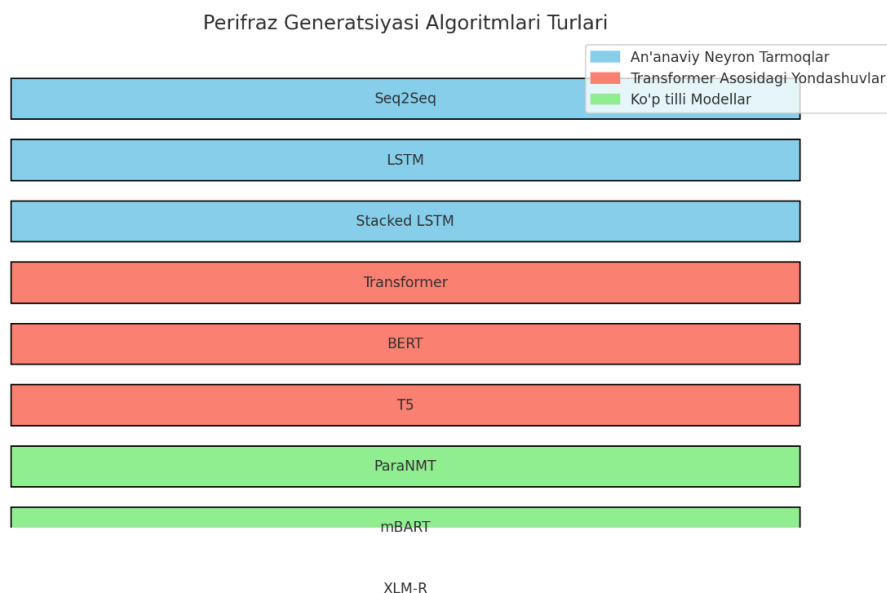
Seq2Seq (sequence-to-sequence) arxitekturasida perifrasiy generatsiyasida dastlabki samarali yondashuvlardan biri hisoblanadi. Ushbu modelning asosiy tamoyili - berilgan kirish matnini (input sentence) kodlovchi (encoder) blok orqali yashirin holatda ifodalab, uni chiqish matniga (output sentence) dekodlovchi (decoder) blok orqali qayta ishlashdir. Ayniqsa, LSTM (Long Short-Term Memory) hujayralari uzoq muddatli bog'lanishlarni saqlay olish qobiliyati bilan bu modelga kuchli kontekst anglash imkonini beradi. Biroq ushbu arxitektura ketma-ketlikka asoslanganligi sababli, parallel ishlov berish imkoniyati cheklangan bo'lib, bu esa katta hajmdagi matnlar bilan ishlashda hisoblash xarajatlarini oshiradi. Bundan tashqari, LSTM modellar murakkab sintaktik yoki semantik o'zgarishlarni chuqur anglay olishda ko'pincha qiyinchiliklarga duch keladi [7].

2017-yilda Vaswani va boshqa tadqiqotchilar [8] tomonidan taklif etilgan Transformer arxitekturasida perifrasiy generatsiyasida yangi bosqichni boshlab berdi. Uning asosiy innovatsiyasi bo'lgan self-attention mexanizmi modelga matndagi har bir tokenning boshqa tokenlar bilan bog'liqligini aniqlash va chuqur kontekstni anglash imkonini beradi. Transformerlarning yana bir katta afzalligi — ular ketma-ketlikni ketma-ket emas, balki parallel ravishda qayta ishlaydi, bu esa o'qitish va inferensiya tezligini sezilarli darajada oshiradi. BERT (Bidirectional Encoder Representations from Transformers) va T5 (Text-to-Text Transfer Transformer) kabi modellarning perifrasiylash uchun moslashtirilgan versiyalari mavjud bo'lib, ular ko'plab benchmark testlarda yuqori natijalarga erishgan. Masalan, T5 modelining "text-to-text" dizayni uni har xil NLP vazifalariga, jumladan, perifrasiy generatsiyasiga moslashtirishni osonlashtiradi [3]. Biroq Transformer modellarining katta hajmi va o'qitish uchun talab etiladigan resurslar ularni cheklangan kompyuter imkoniyatlariga ega tadqiqotchilar uchun qiyinlashtiradi.

So'nggi yillarda ko'p tilli (multilingual) modellar perifrasiy generatsiyasida katta rol o'ynay boshladi. Ayniqsa, past resursli tillar uchun bu yondashuvlar katta imkoniyatlar yaratmoqda. ParaNMT-50M korpusi yordamida o'qitilgan modellar turli tillar oralig'ida o'xshash ma'nodagi jummalarni generatsiya qilish imkonini beradi [10]. Bundan tashqari, mBART (Multilingual BART) va XLM-R (Cross-lingual RoBERTa) modellarining ko'p tilli korpuslarda o'qitilganligi ularni o'zbek tiliga o'xshash tillar uchun ham foydali qiladi. Bu modellar o'zbek tilida yetarli ma'lumotlar bo'lmasa ham, boshqa turkiy tillarda o'qitilgan kontekstni transfer qilish orqali yaxshi natijalarga erishishi mumkin. Ammo bu modellarning asosiy cheklovi — ular o'zbek tiliga xos bo'lgan grammatik va stilistik xususiyatlarni chuqur tushunmasligi mumkin, chunki ko'p tilli korpuslar orasida o'zbek tilidagi matnlar soni juda kam.

Ushbu texnik tahlil shuni ko'rsatadiki, perifrasiy generatsiyasi uchun eng kuchli yondashuvlar Transformer asosidagi modellar hisoblanadi. Biroq har bir modelning o'ziga xos ishlash doirasi va qo'llanish sohasi mavjud. Past resursli tillar

uchun ayniqsa, ko'p tilli modellar yoki transfer learning asosidagi yondashuvlar yanada samarali bo'lishi mumkin. Shu sababli, algoritm tanlashda til resurslari, vazifa turi va mavjud hisoblash imkoniyatlari e'tiborga olinishi lozim.



1-rasm. Perifraz generatsiyasi uchun qo'llaniladigan asosiy algoritm turlari

Mana bu diagramma perifraz generatsiyasi uchun qo'llaniladigan asosiy algoritm turlarini ko'rsatadi. Ranglar orqali quyidagilar ajratilgan:

- **Ko'k:** An'anaviy neyron tarmoqlar (Seq2Seq, LSTM)
- **Qizil:** Transformer asosidagi yondashuvlar (BERT, T5)
- **Yashil:** Ko'p tilli modellar (mBART, ParaNMT)

Algoritmning taqqoslov tahlili

Quyidagi jadvalda asosiy uch turdagi perifraz generatsiyasi algoritm turlari o'zaro solishtirilgan. Jadvalda har bir yondashuvning kontekst anglash darajasi, hisoblash murakkabligi, resurs talabchanligi va o'zbek tiliga moslashuv darajasi keltirilgan.

1-jadval. Uch turdagi perifraz generatsiyasi algoritm turlarining o'zaro solishtiruv

Algoritm turi	Kontekst anglash	Hisoblash murakkabligi	Resurs talabchanligi	O'zbek tiliga moslashuv
Seq2Seq + LSTM	O'rta	O'rta	Past	Cheklangan
Transformer (BERT, T5)	Yuqori	Yuqori	Yuqori	O'rta
Ko'p tilli modellar (mBART, XLM-R)	Yuqori	Yuqori	Yuqori	O'rta-Yaxshi

Ushbu taqqoslovdan ko‘rinadiki, Transformer asosidagi modellar kontekstni chuqur tushunish qobiliyatiga ega, biroq ularni o‘qitish va ishlatish uchun katta hisoblash resurslari talab etiladi. Ko‘p tilli modellar esa o‘zbek tiliga nisbatan yaxshiroq moslashish imkoniyatiga ega, ayniqsa boshqa turkiy tillar bilan o‘xshashlikdan foydalanilsa.

Perifraz generatsiyasi bugungi kunda ko‘plab real hayotiy vazifalarda samarali qo‘llanilmoqda. Quyida bu algoritmlarning amaliy sohalaridagi ishlatilishiga oid bir nechta misollar keltirilgan: Yevropa va AQShdagi ta’lim texnologiyalari kompaniyalari o‘quvchilar uchun matnlarni avtomatik tarzda soddalashtiruvchi ilovalarda perifraz generatsiyasidan foydalanmoqda. Masalan, murakkab ilmiy yoki texnik matnlar oddiy so‘zlar bilan qayta ifodalanadi, bu esa tushunishni osonlashtiradi. Onlayn o‘quv platformalari foydalanuvchilarga test savollarini avtomatik yaratishda perifrazlashdan foydalanmoqda. Masalan, “OpenAI GPT” asosida ishlovchi vositalar berilgan matndan turli xil formulirovkadagi savollar ishlab chiqadi.

Ba’zi akademik platformalar va ilmiy jurnallar perifraz deteksiyasini amalga oshiruvchi tizimlardan foydalanadi. Reklama va marketing sohalarida bir fikrni bir nechta uslubda ifodalash orqali kreativ kontent yaratiladi. Misol uchun, bir necha variantdagi sarlavhalar, postlar yoki reklama matnlari ishlab chiqilishi mumkin. Perifraz generatsiyasi bir nechta tillarda ekvivalent jummalarni yaratish orqali foydalanuvchiga sinonim variantlar va ifoda shakllarini ko‘rsatadi. Bu yondashuv til o‘rganuvchilarning lug‘aviy boyliklarini oshirishga xizmat qiladi.

O‘zbek tilida perifraz generatsiyasining amaliy jihatlari

Perifraz generatsiyasi algoritmlarini o‘zbek tilida qo‘llashda tilning morfologik boyligi va sintaktik moslashuvchanligi alohida e’tiborni talab qiladi. O‘zbek tili agglyutinativ til bo‘lgani sababli, so‘z yasovchi va qo‘shimcha qo‘shuvchi morfemalarning ko‘pligi perifrazlashda semantik izchillikni saqlab qolish uchun algoritmlardan yuqori darajada kontekstni anglashni talab qiladi.

Masalan, quyidagi jumlani olaylik:

Asl jumla: *Talabalar darsga vaqtida kelishlari kerak.*

Perifraz varianti: *O‘quvchilar mashg‘ulotlarga o‘z vaqtida yetib kelishlari lozim.*

Bu yerda “talabalar” → “o‘quvchilar”, “dars” → “mashg‘ulot”, “kerak” → “lozim” kabi sinonimik almashuvlar mavjud bo‘lsa-da, mazmun o‘zgarmagan. Aynan shunday sintaktik strukturalar asosida qurilgan perifrazlar tilni tushunishga ijobiy ta’sir ko‘rsatadi.

Yana bir misol:

Asl jumla: *Bugun ob-havo iliq va quyoshli bo'ladi.*

Perifraz variantlari:

- *Havo bugun iliq bo'lib, quyosh charaqlab turadi.*
- *Quyoshli va iliq kun bugungi ob-havoni belgilaydi.*

Bu misollar ko'rsatadiki, o'zbek tilida perifrazlash nafaqat so'z darajasida, balki gap tuzilmasi, so'z tartibi va stilistik uslubni o'zgartirish orqali amalga oshiriladi. Shuning uchun, perifraz algoritmlari faqatgina so'z almashtirish emas, balki gap tuzilmalarini qayta qurishga qodir bo'lishi zarur.

O'zbek tilidagi periferik generatsiyani amalga oshirish uchun quyidagi arxitektura tavsiya etiladi:

1. **O'zbek tilidagi annotatsiyalangan kichik korpuslar:** masalan, yangilik matnlari, darsliklar va davlat hujjatlari asosida tuzilgan parallellikdagi variantlar.
2. **mBART modelining transfer learning usuli bilan o'zbek tiliga moslashtirilgan versiyasi.**
3. **Data augmentation (ma'lumotlarni sun'iy ko'paytirish)** orqali sinonimlar, gap tuzilmalari va stilistik o'zgarishlar kiritilgan korpuslar.
4. **BLEU va ROUGE kabi avtomatik metrikalar asosida baholash tizimi.**

Ushbu yondashuv yordamida o'zbek tilidagi perifrazlash vazifalari – masalan, o'quvchilarga soddalashtirilgan matn berish, avtomatik savol yaratish, tilni o'rganishga ko'maklashish va kontent yaratishda takrorlanuvchi fikrlarni stilistik jihatdan boyitish – yuqori sifatda amalga oshiriladi.

Perifraz generatsiyasi algoritmlari zamonaviy tabiiy tilni qayta ishlash sohasida yuqori ahamiyatga ega bo'lgan yo'nalishlardan biri sanaladi. Ushbu maqolada perifrazlash algoritmlarining asosiy turlari, jumladan, an'anaviy seq2seq modellari, Transformer arxitekturasi asosidagi yondashuvlar (BERT, T5) hamda ko'p tilli modellar (mBART, XLM-R) texnik va nazariy jihatdan tahlil qilindi. Shuningdek, o'zbek tilida perifraz generatsiyasi qanday amalga oshirilishi mumkinligi haqidagi muammolar va istiqbolli yondashuvlar ham ko'rib chiqildi.

O'zbek tili o'zining morfologik boyligi va sintaktik moslashuvchanligi bilan boshqa ko'plab tillardan farq qiladi. Shuning uchun bu til uchun perifraz algoritmlarini ishlab chiqishda kontekstni chuqur tushunishga asoslangan yondashuvlar zarur. Maxsus tayyorlangan o'zbek tilidagi korpuslar, sinonimik ma'lumotlar bazalari va stilistik vositalar asosida ishlov beruvchi algoritmlar aynan tilimiz xususiyatlariga moslashtirilgan bo'lishi kerak.

Tahlillar shuni ko'rsatadiki, ayniqsa, Transformer va ko'p tilli modellar yordamida past resursli tillarda ham samarali perifrazlash mumkin. Agar mBART kabi modellar o'zbek tilida qayta o'qitilsa va ma'lumotlar bazasi sun'iy yo'l bilan boyitilsa, tilimizda yuqori sifatli perifrazlash tizimi yaratish mumkin. Masalan, til o'rganish ilovalari, avtomatik tarjima tizimlari, kontent generatsiyasi vositalari va plagiat aniqlash dasturlarida bu yondashuvlar muhim rol o'ynaydi.

Shuningdek, maqolada o'zbek tilidagi real perifraz misollari tahlil qilinib, jumla sathidagi semantik tenglikni saqlab qolish uchun sinonim almashuv, gap tuzilmasini o'zgartirish va stilistik moslashuvlarning qanday ishlashi ko'rsatib o'tildi. Bu jihatlar algoritmik yondashuvlar uchun nazariy asos yaratadi.

Kelgusida o'zbek tiliga xos perifrazlash tizimlarini yaratishda lingvistik resurslar – lug'atlar, frazeologik birliklar, semantik tarmoqlar – asosida modellarni o'qitish va baholash muhim hisoblanadi. Ushbu izlanishlar ilmiy, amaliy va pedagogik sohalarda keng qo'llanilishi mumkin bo'lgan natijalarni taqdim etadi.

Foydalanilgan adabiyotlar:

1. Abdullayev A. O'zbek tilida ekspressivlik ifodalashning sintaksik usullari. – Toshkent, 1983.
2. Rahmatullayev Sh. O'zbek tili frazeologik lug'ati. – Toshkent: G'afur G'ulom nashriyoti, 2022.
3. Turansky I. Sementicheskaya kategoriya intensivnosti v angliyskom yazike. – Moskva: Vysshaya shkola, 1990.
4. Sharipov M. Hozirgi o'zbek tilida so'z birikmalari sintaksisi masalasi. – Toshkent: Fan, 1978.
5. Statista.com. AI usage in education sector worldwide [<https://www.statista.com>] (murojaat sanasi: 2025-04-11).
6. Chen, X., Liu, X., Sun, M., Lin, Y. Paraphrase generation with deep neural networks. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018. (pp. 448–457). [Chen, 2018:448].
7. Prakash, A., Hasan, S. A., Lee, K., Datla, V., Qadir, A., Liu, J., Farri, O. Neural paraphrase generation with stacked residual LSTM networks. In COLING 2016. (pp. 1243–1252).
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Polosukhin, I. Attention is all you need. In Advances in Neural Information Processing Systems, 2017. (pp. 5999–6009).
9. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Liu, P. J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research, 21(140), 1–67. 2020.



10. Mallinson, J., Sennrich, R., Lapata, M. Paraphrasing revisited with neural machine translation. In Proceedings of the 15th Conference of the European Chapter of the ACL, 2017). (pp. 400–410). [Mallinson et al., 2017:400].
11. Çelikyilmaz, A., Bosselut, A., He, X., Choi, Y. Generating story endings with planning and style-aware reward. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics, 2021.
12. Al-Shameri, N., Al-Khalifa, H. A multilingual study of paraphrase generation models across low-resource languages. Data in Brief, Volume 52, April 2024, 109661. Elsevier, 2024.