



O'ZBEK TILIDAGI GAPLARNI PEREFRAZ QILISH BOSQICHLARI VA MUAMMOLARI

X.X.Axmedova

*Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti
katta o'qituvchisi*

E-mail: khusniya150586@gmail.com

ORCID – 0009-0009-7804-6395

Annotatsiya: Ushbu tadqiqot perefrazlar taksonomiyasi tushunchasi va uning leksik, sintaktik hamda semantik darajalardagi ierarxik tuzilishi batafsil yoritilgan. Semantik perefrazlar esa mazmunni umumlashtirish, aniqlashtirish, tafsilot qo'shish yoki yashirin qarama-qarshilik yaratish asosida tahlil qilingan. Shuningdek, perefraz yaratishning an'anaviy (sinonimlarni almashtirish, WordNet/tezaurusga asoslangan) va neyron modellar (Seq2Seq, LSTM) asosidagi usullari hamda ularning afzallik va cheklovlari ko'rib chiqilgan. O'zbek tilidagi so'z, ibora va gap darajasidagi perefrazlash imkoniyatlari misollar orqali yoritilib, perefraz generatsiyasi NLP ilovalari - savol-javob tizimlari, mashina tarjimai va axborot izlashda muhim ahamiyatga ega ekani asoslab berilgan.

Kalit so'zlar: Perefraz, taksonomiya, NLP, sinonimlar, semantika, neyron modellar.

Abstract: This study presents a detailed discussion of the concept of paraphrase taxonomy and its hierarchical structure based on lexical, syntactic, and semantic levels. Semantic paraphrases are analyzed in terms of meaning generalization, specification, addition or removal of details, and the creation of explicit or implicit contrast. In addition, both traditional approaches to paraphrase generation (such as synonym replacement based on WordNet or thesauri) and neural model-based methods (Seq2Seq, LSTM) are reviewed, along with their advantages and limitations. The capabilities of paraphrasing at the word, phrase, and sentence levels in the Uzbek language are demonstrated through examples, and the significance of paraphrase generation for NLP applications-such as question answering systems, machine translation, and information retrieval-is substantiated.

Keywords: Paraphrase, taxonomy, NLP, synonyms, semantics, neural models.

Аннотация: В данной статье подробно рассматривается понятие таксономии перефразирования и её иерархическая структура, основанная на лексическом, синтаксическом и семантическом уровнях. Семантические перефразы анализируются с точки зрения обобщения и уточнения смысла, добавления или удаления деталей, а также создания явного или скрытого противопоставления. Кроме того, рассматриваются как традиционные подходы к генерации перефраз (такие как замена синонимов на основе WordNet или тезаурусов), так и методы, основанные на нейронных моделях (Seq2Seq, LSTM), с анализом их преимуществ и ограничений. Возможности перефразирования на уровне слова, словосочетания и предложения в





узбекском языке продемонстрированы на примерах, а также обоснована значимость генерации перефраз для приложений NLP, таких как системы вопросно-ответного взаимодействия, машинный перевод и информационный поиск.

Ключевые слова: перефразирование, таксономия, NLP, синонимы, семантика, нейронные модели.

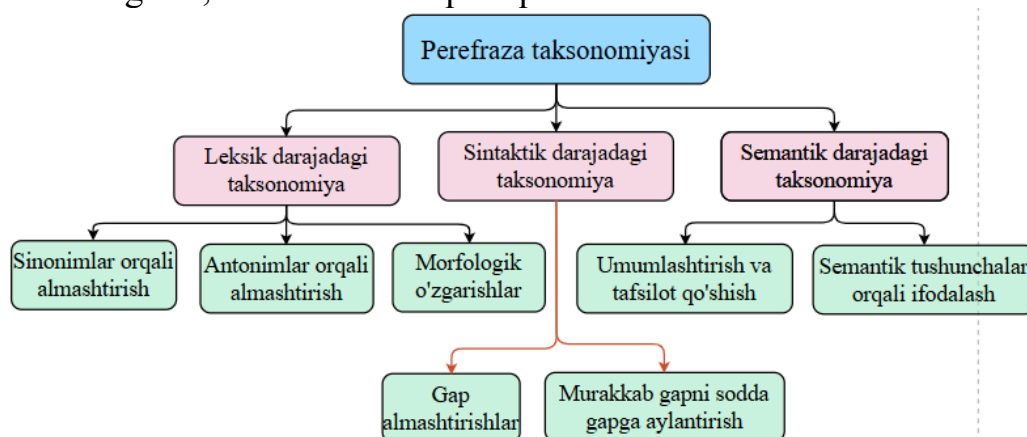
KIRISH

Tabiiy tilni avtomatik qayta ishlash muammosi yarim asrdan ortiq vaqt davomida ilmiy tadqiqotlarning muhim yo'nalishlaridan biri bo'lib kelmoqda. Mazkur muammoning murakkabligi tabiiy tillarning ko'p qirraliligi va ularni to'liq tavsiflab beruvchi yagona nazariy modelning mavjud emasligi bilan izohlanadi. Shu sababli, matn va gaplarga avtomatik ishlov berishda lingvistik analizatorlardan foydalanish muhim ahamiyat kasb etadi.

Semantik tahlil tabiiy tilga ishlov berish jarayonida semantik qidiruvni amalga oshirish imkonini beradi. Semantik tahlil qanchalik mukammal bo'lsa, qidiruv natijalari shunchalik samarali va aniq bo'ladi. Semantik tahlilning samaradorligi bevosita lingvistik resurslarga bog'liq. Bular orasida leksik resurslar - lug'atlar, tezauruslar va ontologiyalar - muhim rol o'ynaydi. Lug'at esa leksik axborotning eng an'anaviy shakli bo'lib, u leksik birliklar (odatda so'z yoki so'z birikmasi), ularning tuzilishi va qamroviga qarab farqlanadi. Lug'atlar turlicha bo'lishi mumkin: aniq sohalar atamalarini qamrab oluvchi lug'atlar, umumiy leksika lug'atlari, shuningdek, sinonimlar va paronimlar lug'atlari.

MATERIALLAR VA USULLAR

Taksonomiya - bu obyektlarni, tushunchalarni yoki hodisalarni ma'lum bir tizimli tamoyillar asosida guruhlariga ajratish va ularni ierarxik tuzilishda tartibga solish jarayoni. Perefrazalar taksonomiyasining leksik, sintaktik va semantik darajadagi o'zgarishlarga asoslanishi shuni anglatadiki, jumlar yoki matnlar turli darajalarda o'zgarib, asl ma'noni saqlab qolishi mumkin.



1-rasm: Perefraza taksonomiyasining iyerarxiyasi

Matnlarni perefraz qilish axborot tizimini ishlab chiqishda chuqur taksonomiyadan foydalanish maqsadga muvofiq. Quyida chuqur taksonomiya darajasidan foydalanishni namunalar yordamida keltiramiz.

Leksik perefrazlar (so'z darajasidagi o'zgarishlar):





Leksik birliklarni sinonimik almashtirish:

- Oddiy sinonim (masalan, «tez» → «shiddatli»)
- Kontekstga bog'liq sinonim («ilmiy ish» → «tadqiqot»)
- Metaforik sinonim («ko'zgu» → «haqiqat aks etuvchi narsa»)

Antonimik almashtirish:

- To'g'ridan-to'g'ri antonim («katta» → «kichik»)
- Kontekstual antonim («yaxshi inson» → «yomon odam»)

Morfologik o'zgarishlar:

- So'z turkumining o'zgarishi («yashil» → «yashillik»)
- Fe'l va ot o'rtasida almashtirish («ishonmoq» → «ishonch»)

Qisqartirish va kengaytirish

- Rasman qabul qilingan qisqartirishlar («O'zMU» → «O'zbekiston Milliy Universiteti»)
- Matnni to'liq yoki ixchamroq shaklga keltirish

Sintaktik perefrazlar (gap tuzilishidagi o'zgarishlar)

So'z tartibini o'zgartirish:

- Oddiy («U kitob o'qidi» → «Kitobni u o'qidi»)
- Kuchaytirish uchun o'zgarish («Men sen bilan uchrashishni istayman» → «Seni uchratishni juda ham istayman»)

Murakkab gapni soddalashtirish yoki murakkablashtirish:

- Birikmalarni bo'lish («Biz yo'lda ketayotganda yomg'ir yog'di» → «Biz yo'lda edik. Yomg'ir yog'di.»)
- Soddalashtirish («Men bu kitobni o'qiganimdan so'ng, yangi narsalarni tushundim» → «Bu kitob menga yangi narsalarni tushuntirdi»)

Xronologik tartibni o'zgartirish

- Oldin-ketin munosabatlarini o'zgartirish («Dars tugadi, keyin biz uyga ketdik» → «Biz uyga keta boshladik, chunki dars tugagan edi.»)

Semantik perefrazlar (mazmun va kontekst darajasidagi o'zgarishlar):

Tafsilot qo'shish yoki olib tashlash

- Qo'shish («U kech keldi» → «U tirbandlik sababli kech keldi»)
- Olib tashlash («Ishga kechikkanim sababli jazo oldim» → «Men jazo oldim»)

Mazmuniy umumlashtirish yoki aniqroq ifodalash

- Umumlashtirish («U bog'da yurdi, gul hidladi va quyosh botishini tomosha qildi» → «U bog'da dam oldi.»)
- Aniqroq ifoda («U hozir shunday ahvolda» → «U hozir juda tushkun kayfiyatda.»)

Qarama-qarshi fikr yaratish

- To'g'ridan-to'g'ri («U baxtiyor edi» → «U baxtsiz edi»)
- Yashirin qarama-qarshilik («Uning yuragida quvonch bor edi» → «U tashqaridan quvnoq ko'rindi, lekin ichida qayg'u bor edi.»)

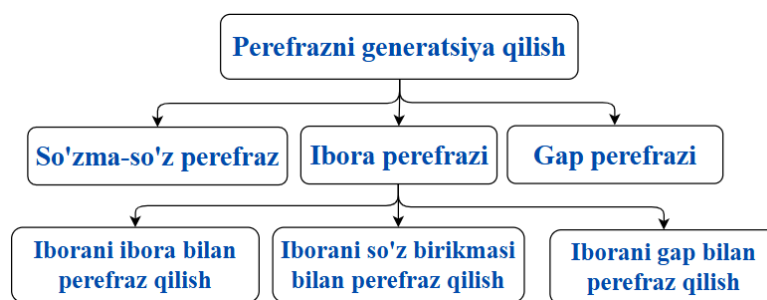




Muhokama va tahlil

Tabiiy tilga ishlov berish sohasida gaplarni perefraz qilish, gap ekvivalentlarni yaratish, gaplarni semantik generatsiya qilish tushunchalari o'zaro sinonim tushunchalardir. Gapning perefrazi uning ma'nosini saqlab qoladi, biroq asl shakldan farqli so'z va iboralarni tanlaydi. Perefraz yaratish (Paraphrase Generation-PG) turli tabiiy tilga ishlov berish (Natural Language Processing – NLP) ilovalari ishlab chiqishda muhim masala bo'lib, u tabiiy til birliklarini kengaytiradi. Perefraz yaratish *savol-javob tizimlari*, *mashina tarjimasi* va *axborot izlash* kabi ko'plab vazifalarni bajarishda uchun muhim rol o'ynaydi. Savol - javob tizimlarida PG semantik jihatdan ekvivalent savollarni hamda ularga turli ko'rinishdagi javoblarni yaratish uchun ishlatilishi mumkin. Bu javoblarni qidirishda so'zlar mos kelmasligi muammolarini hal qilishga yordam beradi [1]. Ilmiy ishlar, badiiy asarlardagi semantik jihatdan takrorlanuvchi jumalarni aniqlash orqali matnlarni annotatsiyalashga yordam beradi. Bundan tashqari, mashina tarjimasi uchun perefrazlash, ma'lumotlar bazasining kamligini bartaraf etishga yordam berishi mumkin, chunki u ma'lumotni kengaytiradi [2].

Tabiiy tillarda til birliklarini perefraz qilishni quyidagicha amalga oshirish mumkin:



2-rasm: Perefrazni yaratish usullari

So'zma - so'z perefraz qilishda o'zbek tilidagi gap tarkibida mavjud bo'lgan leksik birliklarni sinonimlari bilan almashtirish natijasida hosil qilinadi. Bunda har bir leksik birlikning sinset to'plami miqdorining kombinatsiyalari ishtirokida turlicha gaplar tuzish mumkin.

1-jadval: Gapni leksik perefraz qilish namunalari

1	Ertalab badantarbiya bilan shug'ullangan bolalar sog'lom bo'ladi.
2	Tongda badantarbiya bilan shug'ullangan bolalar sog'lom bo'ladi.
3	Ertalab badantarbiya qilgan bolalar sog'lom bo'ladi.
4	Ertalab sport bilan shug'ullangan bolalar sog'lom bo'ladi.
5	Tongda sport bilan shug'ullangan bolalar sog'lom bo'ladi.
6	Ertalab sport mashg'ulotlari bilan shug'ullangan bolalar sog'lom bo'ladi.
7	Ertalab sport mashg'ulotlari bilan shug'ullangan bolalar kasal bo'lmaydi.

1-jadvaldan ma'lumki leksik perefraz qilishda har bir so'zning sinonimlarini qo'llash orqali turli gaplar hosil qilish mumkin. Masalan ertalab so'zini turli so'zlar bilan almashtirish mumkin.





Ertalab soʻzi uchun quyidagi sinonimlarni keltirish mumkin:

Tong, Sahar, Subh, Tong Sahar, Saharda, Tongda, Sahar payti, Subhida

Badantarbiya soʻzining sinonimlari: Jismoniy tarbiya

Sport mashgʻulotlari, Tanani chiniqtirish, Jismoniy mashqlar, Gimnastika

Mashq, Sport

Soʻzlarning antonimlarini uni inkor qilish orqali gapning semantikasini saqlab qolish mumkin. Yaʼni

Ertalab sport mashgʻulotlari bilan shugʻullangan bolalar kasal boʻlmaydi.

Sogʻlom boʻladi brikmasini kasal boʻlmaydi birikmasi bilan almashtirish gapning maʼnosiga taʼsir qilmaydi. Bitta *sogʻlom boʻlmoq* birikmasini antonimlarini qoʻllash orqali quyidagi gaplarni hosil qilish mumkin.

1. *Ertalab badantarbiya bilan shugʻullangan bolalar **kasal boʻlmaydi.***
2. *Ertalab badantarbiya bilan shugʻullangan bolalar **zaiflashmaydi.***
3. *Ertalab badantarbiya bilan shugʻullangan bolalar **betob boʻlmaydi.***
4. *Ertalab badantarbiya bilan shugʻullangan bolalar **holsizlikka uchramaydi.***
5. *Ertalab badantarbiya bilan shugʻullangan bolalar **xastalikka chalinmaydi.***

Gapdagi soʻzlarning antonimlari yordamida perefrazlashda gapdagi kesimning inkor shakli ishlatiladi. Yuqoridagilardan koʻrinadiki, leksemalarning sinonim hamda antonimlari yordamida turli perefraz kombinatsiyalar tuzish mumkin. Perefrazlash usuli boʻyicha ilgari bir nechta tadqiqotlar olib borilgan.

Neyron modellar asosida esa bir nechta tadqiqotlar amalga oshirilgan. Prakash, A va boshqalar soʻzlarni sinonimlari bilan perefrazlash uchun Seq2Seq modeliga asoslangan *stacked residual LSTM* tarmogʻidan foydalanganlar [3]. Ushbu yondashuv tabiiy tilni generatsiya qilishda ancha ilgʻor natijalarni koʻrsatgan. Biroq, boshqa tomondan, hosil boʻlgan perefrazlarning soʻz xilma-xilligi pastligi bu yondashuvning zaif tomonlaridan biri sifatida qayd etilgan [4].

Tabiiy tilga ishlov berish sohasida perefraz qilishning yana bir muhim bosqichi Gap perefrazidir. Bunda gaplarni tarkibida oʻxshash leksemalari mavjud boʻlmasa-da semantik jihatdan aynan teng boʻlgan gap bilan perefraz qilinadi. Hozirgi kunda SPG (Sentence Paraphrase Generation) uchun mashhur usullar, iboralar asosidagi statistik mashina tarjimai texnikalarini qoʻllanilmoqda [5].

Xulosa

Mazkur maqolada perefraz yaratishning anʼanaviy (sinonimlarni almashtirish, tezaurus va WordNet asosidagi yondashuvlar) hamda zamonaviy (neyron tarmoqlar, Seq2Seq va LSTM modellar) usullari tahlil qilindi. Anʼanaviy usullar soddaligi bilan ajralib tursa-da, neyron modellar kontekstni chuqurroq hisobga olishi bilan ustunlikka ega ekani, biroq ayrim hollarda soʻz xilma-xilligining pastligi kabi kamchiliklarga ega ekani taʼkidlandi.

Xulosa qilib aytganda, matnlarni perefraz qilish axborot tizimlarini ishlab chiqishda chuqur va koʻp bosqichli taksonomiyadan foydalanish maqsadga muvofiqdir. Bunday yondashuv perefrazlarni aniqlash va generatsiya qilishda yuqori aniqlikni taʼminlaydi, NLP ilovalari - savol-javob tizimlari, mashina tarjimai, axborot izlash va matnlarni semantik tahlil qilish samaradorligini oshiradi





hamda o'zbek tili uchun boy va moslashuvchan perefranz resurslarini shakllantirishga zamin yaratadi.

Foydalanilgan adabiyotlar

1. J Edward Hu, Rachel Rudinger, Matt Post, and Benjamin Van Durme. Parabank: Monolingual bi text generation and sentential paraphrasing via lexically constrained neural machine translation. arXiv preprint, 2019.
2. Lin, D., Pantel, P.: Discovery of inference rules for question-answering. Natural Language Engineering 7(4), 343–360 (2001)
3. Zhou, J., & Bhat, S. (2021). Paraphrase Generation: a survey of the state of the art. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing 5075-5086-p. <https://doi.org/10.18653/v1/2021.emnlp-main.414>
4. Prakash, A., Hasan, S. A., Lee, K., Datla, V. V., Qadir, A., Liu, J., & Farri, O. (2016). Neural Paraphrase Generation with Stacked Residual LSTM Networks. International Conference on Computational Linguistics, 2923–2934. Retrieved from <https://www.aclweb.org/anthology/C16-1275.pdf>
5. Lin, Z., & Wan, X. (2021). Pushing Paraphrase Away from Original Sentence: A Multi-Round Paraphrase Generation Approach. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 1548-1557-p. <https://doi.org/10.18653/v1/2021.findings-acl.135>

