

SINTAKTIK SINONIMIYA VA UNI ANIQLASH XUSUSIDA

Uzoqova Mohiyaxon Tuyg'un qizi

Kompyuter lingvistikasi mutaxassisligi magistranti

Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti.

Annotatsiya: Sinonimlar turli belgilariga ko'ra, shu jumladan, mazmuni va shakliga ko'ra tasniflanishi mumkin. Shakliy sinonimlar qatoriga sintaktik sinonimlarni ham kiritish mumkin. Ushbu maqolada sintaktik sinonimiya hodisasiga va uning shakllanish usullariga to'xtab o'tilgan. Shuningdek, sintaktik sinonimiyaning faqat algoritmik yo'l bilan avtomatik aniqlash murakkabligi bois, jarayonni soddalashtiruvchi va muallif tomonidan samaraliroq deb baholanayotgan til modellaridan biri – T5 modeliga ham qisqacha to'xtalib o'tilgan.

Kalit so'zlar: Sinonimlar tasnifi, sintaktik sinonimiya, til modellari, T5 modeli, parafrazlash.

Abstract: This paper explores the complex realm of synonymy, focusing on its various classifications based on meaning and form, including syntactic synonymy. It delves into the intricate methods of shaping syntactic synonymy, acknowledging the challenges in automatic identification using algorithms alone. The paper proposes the use of advanced AI models, specifically the T5 language model, as a more effective solution for this linguistic task. The study aims to enhance our understanding of synonymy and contribute to the development of more sophisticated language processing tools.

Keywords: classification of synonyms, syntactic synonymy, language models, T5 model, paraphrasing.

Аннотация: В данной статье исследуется сложный мир синонимии, уделяя особое внимание ее различным классификациям, основанным на значении и форме, включая синтаксическую синонимию. Он углубляется в сложные методы формирования синтаксической синонимии, признавая проблемы автоматической идентификации с использованием одних только алгоритмов. В статье предлагается использовать передовые модели искусственного интеллекта, в частности языковую модель T5, как более эффективное решение этой лингвистической задачи. Исследование направлено на улучшение нашего понимания синонимии и содействие разработке более сложных инструментов языковой обработки.

Ключевые слова: классификация синонимов, синтаксический синоним, модели языка, модель T5, перефразирование.

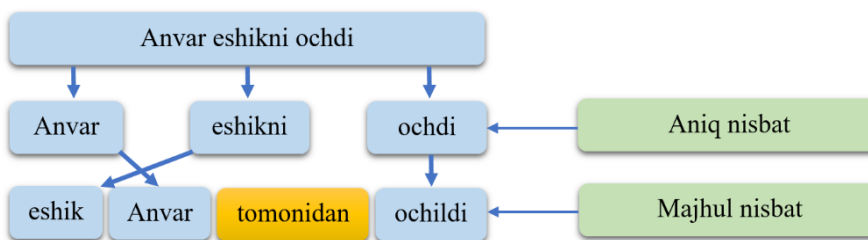
Sinonimlarni tasniflashda turli yondashuvlar mavjud. Umumiy holatda ularni mazmuniy va shakliy sinonimlar kabi yirik guruhlarga bo'lish mumkin. Mavjud tasniflarni umumlashtirgan holda biz shakliga ko'ra tasniflanuvchi sinonimlar sifatida leksik, morfologik, paremeologik va sintaktik sinonimlar taklif qilamiz. Sintaktik sinonimlarning mavjudligi haqida o'zbek tilshunosligida Pinxasov D. [Пинхасов, 1969; 16] va Polvonovaning B. [Polvonova, 2021; 891] ham qayd etib o'tgan.

Sintaktik sinonimlar – bu gapning mazmuni saqlangan, ammo sintaktik o‘zgarish kuzatilgani sababli gapning shakli o‘zgaragan sinonimlardir. Sintaktik sinonimlikni ba’zi olimlar *sintaktik variatsiyalar* deb ham atashadi [Rashidli, 2022; 74].

Sintaktik sinonimlikni hosil qilishning bir qancha usullari mavjud:

1. Gapning nisbatini o‘zgartirish orqali.

Gapning nisbatini o‘zgartirish orqali uning mazmunini saqlagan holda umumiy shaklini o‘zgartirish mumkin. Bunday holatda sintaktik sinonimlik ko‘pincha aniq nisbat – majhul nisbat o‘rtasida kuzatiladi (Qarang: 1 -rasm).



1-rasm. Gapning nisbatini o‘zgartirish orqali sintaktik sinonimlikni hosil qilish

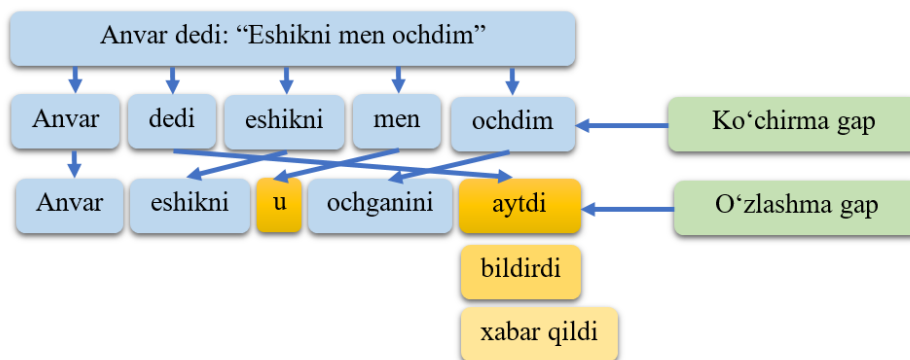
2. Ko‘chirma gapni o‘zlashtirma gapga o‘zgartirish orqali.

Ko‘chirma gapda o‘zga gapining asosiy mazmuni aks etadi, muallif gapida esa ko‘chirma gap kimga tegishli ekanligi, unga xos xususiyat, qanday maqsad va tarzda aytilgani va boshqa qo‘shimcha tavsifiy ma’lumotlar beriladi. Sintaktik sinonimlikni ayan ko‘chirma gapni o‘zlashtirma gapga aylantirish yoki aksini qilish orqali ham hosil qilish mumkin. Bunda o‘zgaralar gapining mazmuni saqlanib, grammatik shakli yoki gap qurilishi o‘zgartirib yuboriladi (Qarang: 2 -rasm):

– Ko‘chirma gapni o‘zlashtirma gapga aylantirishda ko‘chirma gapdagi faol qo‘llanuvchi demoq fe’li aytmoq, gapirmoq, ta’kidlamoq fe’llariga almashtiriladi;

– Ko‘chirma gap o‘zlashtirma gapga aylantirilganda, ba’zan uning uslubiy xususiyatlari ham o‘zgaradi. Ayniqsa, so‘zlovchining nutq jarayoniga turli munosabatini ifodalovchi ayrim kirish so‘zlar, undovlar; *axir, xo‘sh, ha, yo‘q, xo‘p, mayli* kabi gap bilan bog‘lanmagan elementlar, so‘zgaplar ko‘pincha tushib qoladi.

– Undov gap, atov gap, ritorik so‘roq gaplar o‘zlashtirma gapgdan tushirib qoldiriladi.



2-rasm. Gapning nisbatini o‘zgartirish orqali sintaktik sinonimlikni hosil qilish

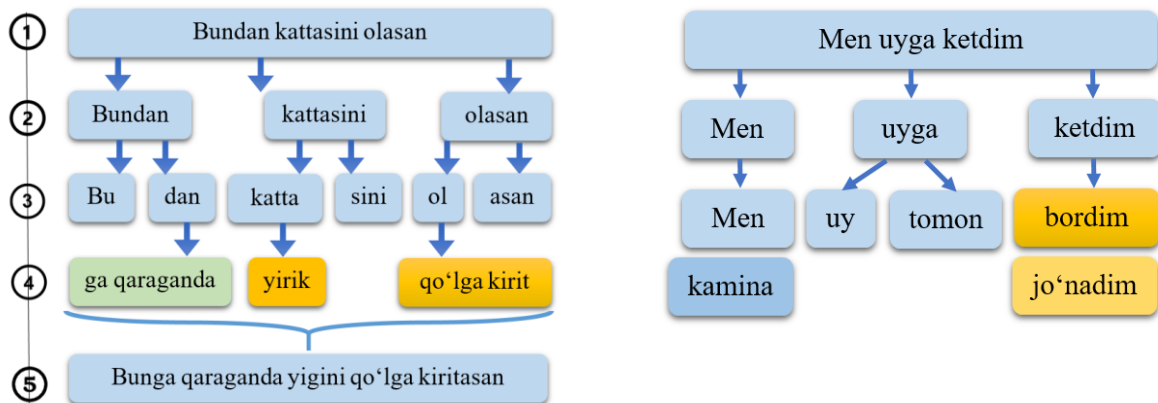
Biroq ko‘chirma gapni o‘zlashtirma gapga aylantirish orqali sintaktik sinonimlikni hosil qilish, bizningcha, sintaktik sinonimizatsiyadagi eng qiyin jarayonlardandir. Chunki ko‘chirma gap qanchalik murakkab tuzilgan bo‘lsa, uni o‘zlashtirma gapga aylantirish shunchalik qiyinlashadi. Bu murakkab sintaktik parsing qilish jarayonini talab qiladi. Bizning ilmiy tadqiqotimiz doirasida matnni sintaktik analiz qilish vazifasi qo‘yilmagani bois bunday sinonimlashni o‘zbekcha sintaktik analizator ishlab chiqilgach unga integratsiya qilish orqali amalga oshirishni istiqbolli reja sifatida belgilab oldik.

3. So‘z yuklamalarning shakl yasovchi qo‘shimchalar bilan o‘zgarishi.

Berilgan gap so‘zlar tarkibidagi ayrim shakl yasovchi qo‘shimchalarning “qo‘shimcha + so‘z yuklama”larga almashishi (Qarang: 3-rasm) yoki so‘z yuklamalarga (Qarang: 4-rasm) ham shakliy o‘zgarishga dalolat qiladi.

4. Tasviriy ifoda yoki izohlovchilar yordamida gapni o‘zgartirish orqali.

Tasviriy ifoda yoki izohlovchilar ham o‘z doirasida (tasviriy ifoda – tasviriy ifoda; izohlovchi – izohlovchi), ham o‘zaro perependikulyar (tasviriy ifoda – izohlovchi; izohlovchi – tasviriy ifoda), ham leksema bilan almashib kelishi mumkin (Qarang: 5-rasm). Bunda gapning sezilarli qismi shaklan o‘zgarishi hisobiga gap qurilishi ham o‘zgarishi mumkin.



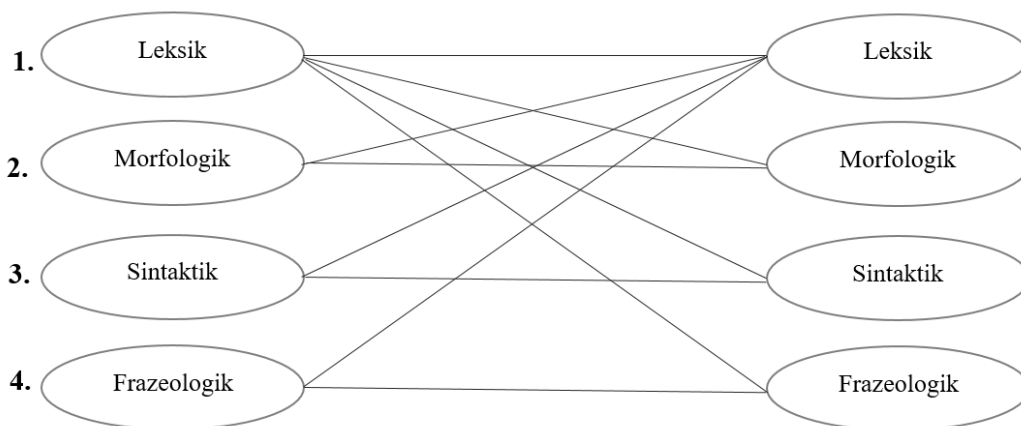
3-rasm. Gapning nisbatini o‘zgartirish orqali sintaktik sinonimlikni hosil qilish

4-rasm. Gapning nisbatini o‘zgartirish orqali sintaktik sinonimlikni hosil qilish

Sintaktik sinonimlar, o‘z navbatida, matn ichida boshqa sinonimik birliklar bilan kesishishi mumkin.

Sintaktik sinonimlarning boshqa turdagi sinonimlar bilan kesishishi:

1. Sintaktik sinonim – leksik sinonim: Toshkent – O‘zbekistonning poytaxti.
2. Sintaktik sinonim – sintaktik sinonim: Aziz eshikni ochdi – Eshik Aziz tomonidan ochildi.
3. Sintaktik sinonim – sintaktik sinonim: “Darsga kim javob beradi?” – dedi ustoz – Ustoz darsga kim javob berishini so‘radi.
4. Sintaktik sinonim – sintaktik sinonim: O‘zbekistonning poytaxtiga xush kelibsiz – Yurtimizning yuragiga xush kelibsiz.



5-rasm. Sinov uchun berilgan gap bo‘yicha asosida modelning kichik dataset bilan mashq qildirilgan dastlabki versiyasidan olingan eng yuqori beshta natija

Sintaktik sinonimlarning bunday turli sinonimlar bilan ham, o‘z turida ham matnda sinonimlasha olishi algortmik yo‘l bilan avtomatik tarzda sintaktik sinonimizatsiyani amalga oshirishni murakkablashtiradi. Buning uchun yoki kuchli lingvistik va texnik tizim asosida ishlovchi morfoanalizator va sintaktik analizatorlarni sinonimayzer dasturiga integratsiya qilish lozim yoki matnni parafaraza qilishga yo‘naltirilgan sun‘iy intellekt modellariga murojaat qilish lozim.

Ana shunday modellardan biri sifatida biz keyinchalik tadqiqotimizga joriy qilishni ko‘zlayotgan T5 modelini taklif qilmoqchimiz [Huggingface, 4]. T5 yoki Text-to-Text Transfer Transformer universal til modeli NLPning juda ko‘p masalalarini hal qilishda: mashina tarjimasida [Huggingface, 5], matnni klassifikatsiya qilishda [Huggingface, 6], NER (named entity recognition) aniqlashda [Huggingface, 7], matnni avtomatik generatsiya qilishda [Huggingface, 8], matnni xulosalashda [Huggingface, 9], o‘rindosh taklif qilishda (Fill-Mask) [Huggingface, 10], savol-javob tizimlarini qurishda [Huggingface, 11], shuningdek, matnni parafraza qilishda foydalaniladi. Demak, bu modeldan gapni avtomatik parafrazalashda ham foydalanish mumkin.

Buning uchun quyidagi ketma-ketlikda ish bajarish lozim:

1. Kerakli datani mos formatda to‘plash.
2. T5 modelini sozlash yoki avvaldan mashq qildirilgan hamda sozlangan T5 modelini topish.
3. Kiruvchi va chiquvchi ma’lumotlarni taqdim etish.
4. Natijani sifat nazoratidan o‘tkazish va zarur bo‘lsa, modelni mukammallashtirish.

Modelni mashq qildirish uchun o‘zbek tilidagi kerakli formatdagi ma’lumotlar korpusini to‘plash quyidagicha ko‘rinishda bo‘lishi lozim:

- 1. Data to‘plash boqichi.** Buning uchun qulay kengaytma bilan saqlangan (masalan, .xsl) “Almashuvchi” va “almashtiriluvchi” qatorlar jufligini hosil qilish kerak (Qarang: 1- jadval).

Almashuvchi gap (“Input”)	Almashtiriluvchi gap (“Target”)
Odatda o‘quvchilarimga shu kundaligimni namuna qilib ko‘rsataman.	Ko‘p hollarda bolalarga shu kundalik namuna qilib ko‘rsatilar edi.
Kundalikni qo‘lma-qo‘l olib tomosha qila boshlashdi.	Kundalikni qo‘llaridan qo‘ymay tomosha qilishdi.

1-jadval. Sinov uchun berilgan gap bo‘yicha asosida modelning kichik dataset bilan mashq qildirilgan dastlabki versiyasidan olingan eng yuqori beshta natija

2. To‘plangan datani tayyorlash bosqichi. Bu bosqichda tayyorlangan juftliklarni tozalash, tokenizatsiya qilish va model tushunishi mumkin bo‘lgan formatga solish lozim.

3. Tayyor datani qismlarga bo‘lish bosqichi. Bu bosqichda tayyor datani 3 qismga: mashq qilish uchun, sozlamarni to‘g‘rilash uchun validatsiyaga va testlash uchun bo‘lish lozim.

4. Tayyor datani raqamli ko‘rinishga o‘tkazish (enkodlash) bosqichi.

Shunday qilib, tayyor data asosida modelni mashq qildirish orqali parafrazalash jarayonini yengillashtirish, mukammallashtirish mumkin. Chunki sintaktik sinonimiyaning har bir turi uchun alohida algoritm ishlab chiqish va uni integratsiya qilish katta resurs talab qilish bilan birga, sinonimizatsiya jarayonini sekinlashtirishi mumkin. Parafrazalash modellaridan foydalanish orqali esa bu muammo tezroq, samaraliroq va sifatliroq yechilishi mumkin degan deb hisoblaymiz.

Foydalanilgan adabiyotlar:

1. Пинхасов Д. Ҳозирги ўзбек адабий тили. Лексикология ва фразеология. –Тошкент: Ўқитувчи, 1969. –Б. 16.
2. Polvonova B. O‘zbek tilida sinonimiya talqini. Oriental Renaissance: Innovative, educational, natural and social sciences. VOLUME 1 | ISSUE 11 ISSN 2181-1784, 2021. – B. 891.
3. Vafa Ahmed gizi Rashidli. Syntactic synonymy, syntactic variability, syntactic transformation, syntactic equivalence // “Colloquium-journal” / Philological sciences / 17 (140), 2022, pp. 74-77. <http://doi.org/10.24412/2520-6990-2022-17140-73-77>
4. https://huggingface.co/docs/transformers/model_doc/t5
5. <https://huggingface.co/docs/transformers/tasks/translation>
6. https://colab.research.google.com/github/patil-suraj/exploring-T5/blob/master/t5_fine_tuning.ipynb
7. https://colab.research.google.com/drive/1obr78FY_cBmWY5ODViCmzdY6O1KB65Vc?usp=sharing
8. https://colab.research.google.com/github/NielsRogge/Transformers-Tutorials/blob/master/T5/Fine_tune_CodeT5_for_generating_docstrings_from_Ruby_code.ipynb
9. <https://huggingface.co/docs/transformers/tasks/summarization>
10. <https://github.com/huggingface/transformers/tree/main/examples/flax/language-modeling#t5-like-span-masked-language-modeling>
11. <https://colab.research.google.com/github/snapthat/TF-T5-text-to-text/blob/master/snapthatT5/notebooks/TF-T5-Datasets%20Training.ipynb>