

O‘ZBEK TILIDAGI LISONIY SINTAKTIK QOLIPLARNI NLP USULLARI BILAN
TAHLIL QILISH

Kadirova Ulfatoy Shavkatbek qizi

Andijon davlat universiteti Kompyuter lingvistikasi yo‘nalishi 2-bosqich talabasi

Email: matkarimovaulfatoy@gmail.com

Annotatsiya

Bugungi globallashtirish jarayonida tabiiy tilni kompyuter yordamida tushunish va qayta ishlash muammosi sun‘iy intellekt sohasidagi eng muhim yo‘nalishlardan biriga aylangan. Tabiiy tilni qayta ishlash (Natural Languages Processing, NLP) inson nutqini avtomatik tahlil qilish va tushunishga yo‘naltirilgan fan sohasidir. U informatika va tilshunoslik fanlari bilan kesishib mashinalarga inson tilini samarali tushunish va unga javob qaytarish imkonini beradigan texnologiyalarni ishlab chiqishni maqsad qiladi. Ushbu maqolada o‘zbek tilidagi sintaktik qoliplar va ularni NLP usullari bilan tahlil qilish va aniqlash masalalari yoritiladi.

Kalit so‘zlar: lingvistik qolip, til modellari, sun‘iy intellekt, so‘z birikmasi, gap.

Annotation: *In today's globalization process, the problem of computer-aided understanding and processing of natural language has become one of the most important directions in the field of artificial intelligence. Natural Language Processing (NLP) is a field of science focused on automatic analysis and understanding of human speech. It aims to develop technologies that allow machines to effectively understand and respond to human language by intersecting with the disciplines of Informatics and Linguistics. This article discusses syntactic patterns in the Uzbek language and the issues of their analysis and identification using NLP methods.*

Keywords: *linguistic pattern, language models, artificial intelligence, phrase, sentence.*

Аннотация: *В условиях современного процесса глобализации проблема компьютерного понимания и обработки естественного языка стала одним из важнейших направлений в области искусственного интеллекта. Обработка естественного языка (NLP) – это область науки, ориентированная на автоматический анализ и понимание человеческой речи. Она направлена на разработку технологий, позволяющих машинам эффективно понимать человеческий язык и реагировать на него, путем взаимодействия с дисциплинами информатики и лингвистики. В данной статье рассматриваются синтаксические паттерны в узбекском языке и вопросы их анализа и идентификации с использованием методов NLP.*

Ключевые слова: *лингвистическая модель, языковые модели, искусственный интеллект, фраза, предложение.*

Tilning mukammallashib borishi, rivojlanishi va barqaror yashab qolishi jamiyatga bog‘liq. Tilning o‘ziga xos qonuniyatlari yillar davomida o‘zgarishlarga uchrab shu til doirasidagi lingvistik qoliplarni shakllantiradi. Lisoniy sintaktik qoliplar – bu til birliklarining jamiyatda barqaror shakllangan ko‘rinishidir. Ular odamlarning so‘z birikmalarini erkin tuzish malakasi asosida asta-sekin rivojlanadi. Bu jarayon uzoq vaqt davomida shakllanib, ijtimoiy tajriba orqali mustahkamlanadi. Natijada, ushbu qoliplar jamiyat a‘zolarining ongida o‘rnashib qolgan bo‘lib, kundalik muloqotda faol qo‘llaniladi.

Sintaktik qolip bu soʻz birikmasi yoki gap tarkibidagi soʻzlarning grammatik munosabatlari hamda ularning ketma-ketlik tartibini aks ettiruvchi grammatik strukturadir. Masalan, “Men maktabga bordim” gapida “Men”- ega (S- Subject), “maktabga” – toʻldiruvchi (O- Object), “bordim” – kesim (V-Verb) tuzilgan boʻlib soʻz tartibi SOVdir va bu oʻzbek tiliga xos soʻz tartibi. Sintaktik qoliplar tilning turli tuzilmalarida namoyon boʻlib, sintaktik qolip avtomatik tarjima, matnni tahlil qilish kabi NLP vazifalarini bajarishda asosiy rol oʻynaydi.

Oʻzbek tilining morfosintaktik qurilishi uni NLP usullari bilan tahlil qilishda oʻziga xos yondashuvni talab qiladi. Birinchidan, tilimizning agglyutinativ tabiati tufayli bitta soʻz shakli ichida soʻz oʻzgartiruvchi va soʻz yasovchi qoʻshimchalarning ketma-ket kelishi murakkab sintaktik munosabatlarni aks ettiradi. oʻqi + t + uvchi + lar + dan soʻzida “oʻqi” asosli feʼlga -t feʼl yasovchi, -uvchi ot yasovchi, -lar koʻplik yasovchi qoʻshimchalar va chiqish kelishigi kelishi bunga yaqqol misol boʻla oladi.

Ikkinchidan, oʻzbek tilida soʻz tartibining nisbatan erkinligi sintaktik qoliplarni aniqlashda qoʻshimcha qiyinchiliklar tugʻdiradi. Masalan, “Men maktabga bordim” gapida ega, toʻldiruvchi va kesimning gapdagi oʻrni oʻzgarishi (S-O-V dan O-S-V yoki S-V-Oga) mantiqiy urgʻuni oʻzgartirsa-da, grammatik bogʻlanishni saqlab qoladi. Yaʼni “Men maktabga bordim” gapi “Maktabga men bordim”, “Men bordim maktabga” gaplari bilan semantik mazmun oʻzgarmasligini taʼminlaydi.

Uchinchidan, oʻzbek tilida omonim va polisemantik soʻzlar koʻplab uchraydi. “yoz” soʻzi “fasl” va “yoz” harakatning bir turini ifodalaydi yoki “oʻt” soʻzi esa “olov”, “maysa” va “oʻt” harakat buyrugʻi boʻlib keladi.

Shu bois, zamonaviy NLP tizimlarida faqat soʻzlarning ketma-ketligiga emas, balki ular orasidagi tobelik munosabatlarini ifodalovchi ierarxik tuzilmalarga asoslanish tilning tabiiy mantigʻini toʻgʻri anglash imkonini beradi.

NLP sohasida sintaktik tahlilni amalga oshirish uchun avvalo tilning morfologik, sintaktik va semantik xususiyatlarini chuqur oʻrganish darkor. Sintaktik qoliplarni aniqlash uchun turli yondashuvlar mavjud boʻlib, ulardan qoidalarga asoslangan (Rule-based) usul – tilshunoslar tomonidan yozilgan grammatik qoidalar asosida gap tuzilmalari tahlil qilinibdir. Bugungi kunda oʻzbek tilida soʻz birikmasining 18 ta umumlashmasi – lisoniy sintaktik qolipi mavjud boʻlib,¹ bu qoliplarning har biridan cheksiz soʻz birikmalari hosil boʻladi. Quyida ushbu qoliplardan bir nechta keltirilgan:

[ot qaratqich kelishigi + ot egalik qoʻshimchasi] – maktabning bogʻi, olmaning poʻchogʻi;

[ot tushum kelishigi+ feʼl]- koʻchani supurmoq, uyni tozalamoq;

[ot qaratqich kelishigi + sifat egalik qoʻshimchasi]- insonning yomoni, ovqatning shirini;

[ot joʻnalish kelishigi + feʼl] – maktabga bormoq, qishloqqa jonamoq;

[ot chiqish kelishigi + feʼl] – daraxtdan yiqildi, tomdan tushdi;

[sifat + ot] – ogʻir yuk, yumshoq mato;

[son + ot] – besh bola, olti soat;

[ravish + feʼl] – tez yurdi, sekin gapirdi.

Gapning sintaktik qoliplari va ularning hosilalari ham soʻz birikmasidek cheksiz birikmalarni hosil qiladi;

[ega + kesim] – Men keldim. Akam keldi;

¹ Sayfullayeva R. va boshqalar. Hozirgi oʻzbek adabiy tili. – Toshkent: Talqin, 2005. -b.318

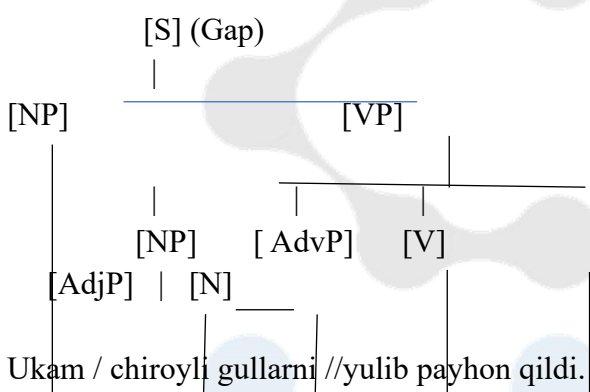
[ega + hol + kesim] – U sekin gapirdi. Men qattiq ishladim;

[ega + to'ldiruvchi + kesim] – Men ovqat tayyorladim.

[ega + aniqlovchi + to'ldiruvchi + kesim] – Men shirin olmani yedim;

[ega + aniqlovchi + to'ldiruvchi + hol + kesim] – Ukam chiroyli gulni yulib payhon qildi.

NLP tizimlari gapdagi so'zlarning ierarxik bog'lanishini tobelik grammatikasi yoki tarkibiy qismlar tahlili orqali amalga oshiradi. Bu jarayonda har bir so'zning gapdagi sintaktik roli va boshqa so'zlar bilan o'zaro aloqasi maxsus 'daraxtsimon' (tree structure) sxemalar yordamida aniqlanadi. Bunday tahlil nafaqat gapning grammatik tuzilishini tushunishga, balki uning leksik-semantik ma'nosini kompyuter tomonidan aniq talqin qilinishiga xizmat qiladi.



Bunday tizimlar aniq va barqaror ishlaydi, lekin yangi va noodatiy holatlarni aniqlashda qiyinchiliklar tug'diradi.

NLP usullaridan yana biri bu statistik yondashuvdir. Bu metod muhim metodlardan biri hisoblanib, til birliklari o'rtasidagi bog'lanishlarni matematik va ehtimollik modellari orqali aniqlaydi. Statistik yondashuv asosan korpus asosida ishlaydi. Korpus- bu ma'lum bir til xilma-xilligini ifodalashga qaratilgan haqiqiy nutq yoki yozuv matning muvozanatli, ko'pincha qatlamli to'plamidir.²

Statistik yondashuvning n-gramm modellarida ma'lum so'zning yuzaga kelish ehtimoli undan avval kelgan so'zlarning ketma-ketligiga bog'liq bo'ladi. n-gramm modellarida unigramm, bigramm va trigramm modellaridan foydalaniladi. Har bir modelning o'ziga xos xususiyati mavjud. Masalan, unigram modelida har bir so'z boshqa so'zlarga taqqoslanmasdan tanlansa, bigramm modeli (ya'ni n = 2) bo'lsa, u holda har bir so'z faqat undan oldingi bitta so'zga qarab tanlanadi. Ya'ni, "men uyga ketyapman" degan jumla unigrammda "men", "uyga", "ketyapman", bigrammda "men uyga" va "uyga ketyapman" bo'lib, "uyga" so'zi "men" so'zidan keyin "ketyapman" so'zi "uyga" so'zidan keyin keladi, model shuni o'rganadi.

Bu usul juda sodda va tez ishlaydi, shuning uchun uni ko'p joyda qo'llashadi. Lekin u uzoq masofadagi grammatik bog'liqliklarni tushunmaydi. Ya'ni, gapning boshidagi so'z bilan oxiridagi so'z orasidagi bog'lanishlarni hisobga olmaydi.

Qisqasi, n-gramm modeli gapdagi so'zlarning ketma-ketligini statistik tarzda o'rganadi, lekin chuqur ma'noni yoki uzoq bog'liqlikni tushunolmaydi.

Statistik yondashuvning yana bir model turi Markov zanjirlari modellarida ham n-gramm modellari kabi gapdagi so'zlarning ehtimolli ketma-ketligi aniqlanadi. n-grammdan farqli ravishda, bu usulda

² https://uz.wikipedia.org/wiki/Korpus_lingvistikasi

ma'lum bir so'z faqat bitta oldingi so'zga qaraydi va so'zlar orasidagi bog'lanish qanday tartibda kelish ehtimolini hisoblab, tilning tuzilishini tushunishga yordam beradi.

Statistik yondashuvda til modellarining takomillashuvi natijasida neyron til modellari paydo bo'lib, avvalgi n-gramm yoki Markov modellaridan farqli ravishda, neyron til modellari ma'lumotlar orasidagi murakkab nisbiy bog'lanishlarni hisobga oladigan bo'ldi. Neyron til modellarining Feedforward Neural Network Language Model (FNNLM) modeli biron so'zni undan oldin kelgan bir nechta so'zga qarab topuvchi neyron model hisoblanadi. Masalan, "men kitob..." deganda model "o'qiyapman" yoki "sotib oldim" degan so'zlarni taxmin qiladi. Ammo bu modelda ham cheklovlar bo'lib, u faqat ma'lum miqdordagi so'zlarni qayta ishlashi mumkin. FNNLM lardan keyin yaratilgan Recurrent Neural Network (RNN) tabiiy tilni qayta ishlashda tilni izchil tahlil qilish imkonini berdi. RNN modellari ketma-ket axborotni bosqichma-bosqich qayta ishlaydi va oldingi so'zlardan olingan ma'lumotni keyingi so'zlarni tahlil qilishda foydalanadi. Bu esa matndagi mantiqiy va semantik bog'lanishlarni yanada aniqroq anglash imkonini yaratadi.

Neyron modellarining eng ilg'or shakli, zamonaviy til modellari – bu transformerlardir.

Transformerlar – bu hozirgi zamon til modellarining (masalan, ChatGPT, BERT, T5 va boshqalar) asosi hisoblanadigan zamonaviy neyron arxitektura. Ular n-gramm yoki Markov zanjirlaridan ancha rivojlangan. Chunki ular butun gap yoki matnni bir vaqtning o'zida ko'radi. Ya'ni, transformer modeli har bir so'zning boshqa barcha so'zlar bilan bog'liqligini hisoblay oladi. Bunga self-attention (o'ziga e'tibor) mexanizmi yordam beradi. Oddiy qilib aytganda, transformer "bu so'z nimaga kelganini" faqat oldingi so'zlarga emas, balki butun matndagi boshqa so'zlarga qarab tushunadi.

Shuning uchun transformerlar uzoq kontekstni ham eslab qoladi, ma'noni chuqur tahlil qiladi va matnni tabiiyroq tushunadi hamda yaratadi. Xullas, transformerlar – tilni "chindan ham tushunadigan" birinchi modellar bo'lgan.

O'zbek tilidagi lisoniy sintaktik qoliplarni NLP usullari yordamida tahlil qilish tilshunoslik va sun'iy intellekt kesishgan muhim tadqiqot yo'nalishi bo'lib, bu ko'p bosqichli, lingvistik tekshiruvlarni talab qiladigan jarayon hamdir. Tilning sintaktik tuzilmalari, ularning o'zaro bog'lanish xususiyatlari va ma'no shakllanishidagi o'rni NLP modellarining aniqligi va samaradorligini bevosita belgilaydi.

Bugungi kunda dunyo miqyosida mashina tarjimasida, matn tahlili, matn generatsiyasi kabi vazifalar aynan sintaktik va semantik tahlilning chuqurligiga tayanadi. Shu bois, o'zbek tilining morfologik, sintaktik va semantik xususiyatlarini hisobga oluvchi modellarni ishlab chiqish dolzarb masala hisoblanadi. Bu borada ushbu yo'nalishda muvaffaqiyatga erishish uchun sifatli korpuslar yaratish, mavjud modellarni o'zbek tiliga moslashtirish va tilimizga xos bo'lgan lingvistik hodisalarni chuqur tahlil qilish zarur. Bu ishlar natijasida o'zbek tiliga mos ilg'or NLP tizimlari – avtomatik tarjimonlar, matn tahlilchilari, til o'rgatuvchi dasturlar va boshqa ko'plab texnologiyalarni yaratish imkoniyati yuzaga keladi.

Foydalanilgan adabiyotlar.

1. *Qurbonova M. va boshqalar. O'zbek tilining struktural sintaksisi. Toshkent: O'zbekiston, 2004. – B.142*
2. *Filipova I. Sun'iy intellektni huquqiy tartibga solish. Samarqand: Samarqand davlat universiteti, 2022. – B.179*
3. *Sayfullayeva R. va boshqalar. Hozirgi o'zbek adabiy tili. Toshkent: Fan va texnologiya, 2009. – B.416*

4. Abdullayeva O. "O'zbek tilida so'z birikmalarining lisoniy sintaktik qoliplari va ularni modellashtirish masalasi"
5. O'zbek tili grammatikasi II tom.
6. Internet manba: <https://mohirdev.uz/blog/Til-modellari-qanday-ishlaydi>
7. Internet manba: https://uz.wikipedia.org/wiki/Korpus_lingvistikasi

