

## O‘ZBEK TILIDAGI SODDA GAPLARNING SINTAKTIK TEGLANGAN BAZASI ORQALI TAHLIL DARAXTINI QURISH

**Ramatova Ma‘mura Gulimboy qizi**

UrDPI Aniq va amaliy fanlar fakulteti o‘qituvchisi

E-mail: [MamuraRahmadova@gmail.com](mailto:MamuraRahmadova@gmail.com)

### ANNOTATSIYA

POS-teglash yoki teqlash -NLP (tabiiy tilni qayta ishlash)ning ishlashning asosiy tushunchalaridan biridir, shuning uchun ham maxsus yaratilgan sintaktik va morfologik teqlash bazasi orqali o‘zbek tilining matnlarini sintaktik tahlil qilishni yaratish o‘zbek tili uchun muhim vazifadir. Ushbu maqolada o‘zbek tili grammatikasi uchun POS-teglardan foydalangan holda o‘zbek tilidagi sodda gaplarning sintaktik tahlil daraxtini qurish haqida so‘z boradi. Maqolamizda maxsus yaratilgan teqlangan baza bundan keyingi ishlarimizda o‘zbek tilidagi sodda gaplarning sintaktik tahlil daraxtini yaratishda muhim ahamiyatga egadir.

**Kalit so‘zlar:** Sintaksis, sintaktik tahlil, sodda gaplar, gap bo‘laklari, NLP, POS-tegging, NLTK.

### ABSTRACT

POS-tagging or tagging is one of the main concepts of NLP (natural language processing), so creating a syntactic analysis of Uzbek language texts through a specially created syntactic and morphological tagging base is an important task for the Uzbek language. For the grammar of the Uzbek language, it is about building a syntactic analysis tree of simple sentences in the Uzbek language using POS-tags. In our article, a specially created tagged base is important in creating a syntactic analysis tree of simple sentences in the Uzbek language.

**Keywords:** Syntax, syntactic analysis, simple sentences, sentence fragments, NLP, POS-tagging, NLTK.

### KIRISH

Bilamizki, matnlarni sintaktik tahlil qilishni modelini ishlab chiqish va uni dasturiy ta‘minotini yaratish uchun bo‘lgan urinishlar XX asrning o‘rtalaridan boshlangan. Sintaktik strukturalar tilshunoslikda amerikalik tilshunos Noam Chomskiyning mashhur asari bo‘lib, dastlab 1957-yilda nashr etilgan. Yuzga yaqin sahifadan iborat qisqa monografiya, Noam Chomskiyning taqdimoti 20-asrning eng muhim tadqiqotlaridan biri sifatida e‘tirof etiladi. Noam Chomskiy o‘z asarida asosan

sintaksisning (gap strukturalarini o'rganish) semantikadan (ma'noni o'rganish) haqida bahs yuritadi. Massachusetts Texnologiya Institutida talabalari uchun tayyorlagan ma'ruza matnlari asosida Xomskiyning birinchi kitobida dastlabki generativ grammatikadagi zamonaviy o'zgarishlar taqdim etildi. Xomskiy bazasida jumllarni kichik qismlarga ajratadigan iboralar tuzilish qoidalaridan foydalanadi. Ular Xomskiy "transformatsiyalar,, deb atagan yangi qoidalar bilan birlashtirilgan. Ushbu protsedura turli xil jumla tuzilmalarini keltirib chiqaradi. (Сергевич.2012.25-bet)

### ASOSIY QISM

Sintaktik kategoriyalar grammatikaning ikkinchi muhim bir bo'lagi bo'lib, morfologik tahlil bilan chambarchas bog'liqdir. Biror gapdagi istalgan gap bo'lagi shakliy jihatdan morfologik shakl va sintaktik shakl qarama-qarshiligi va birligidan iborat. Morfologik kategoriyalar sintaktik shaklning vujudga kelishi uchun asos hisoblanadi. Bir morfologik bo'lak ba'za gaplarda bir necha sintaktik shaklga asos bo'la olganidek, bir necha morfologik shakl bir sintaktik shaklga asos bo'lishi ham mumkin.

Sintaksis bo'limidagi eng muhim aspekt bu-sintaktik tahlildir. Sintaktik tahlilning umumiy maqsadi frazaning funktsional daraxt tuzilmasini yasash, ya'ni, gapning har xil sathda turgan bo'laklari orasidagi bog'liqliklarni aniqlab, ularni shu bo'laklar orasida o'rnatish. Bunday tuzilma formal tillar nazariyasida sintaktik daraxt, chiqarish daraxti yoki frazaning ichki ko'rinishi ham deb ataladi. Frazaning funktsional daraxtini muvaffaqiyatli qurgach, undan ma'naviy elementlarni ajratib olish mumkin hisoblanadi: mantiqiy ob'ekt va sub'ekt, mantiqiy predikat, bevosita va bilvosita to'ldiruvchilarni va hollarni aniqlash mumkin bo'ladi. (Sadullayev, 2013, 46-47-bet)

Sintaktik tahlil yoki tahlil qilish NLP ning uchinchi bosqichidir. Ushbu bosqichning maqsadi aniq ma'noni chizishdir yoki matndan lug'at ma'nosini aytishingiz mumkin. Sintaktik tahlili matnni rasmiy grammatika qoidalarini bilan solishtirganda mazmunlilikini qaysi so'zdan keyin qaysi so'z kelishini avvalroq qaysi so'zlar jumlada kelganini imkon qadar ko'proq o'rganishimizni tekshirishga yordam beradi. Bu ehtiyojni qondirish va qaysi so'zlar avval kelganini tushunish uchun NLP-(Natural Language Processing)da nutq qismlari va sintaktik tahlil qilish juda muhim va ajralmas mavzulardir. (Matlatipov,2009,89-bet)

Grammatik tavsifning eng sodda (primitiv) formal modeli **gap bo'laklariga** ajratish modeli sanaldi. O'zbek tili grammatikasidagi sodda gap ikki qismdan

**-Bosh bo'laklar**

**-Ikkinchi darajali bo'laklar**

Bosh bo'laklar ega va kesimdan tashkil topgan sintaktik qurilma hisoblanadi. Ega nutqning predmetini, kesim esa predmetning harakat belgisini ko'rsatadi. Ega va kesim alohida so'zlar bilan ham, so'zlar guruhi bilan ham ifodalanishi mumkin.

Ega guruhida asosiy mazmun ifodalovchi unsur ot soʻz turkumi bilan ifodalangan ega yoki otlashgan soʻzlar (substantivatsiya), ega guruhidagi boshqa soʻzlar esa eganing aniqlashtiruvchilari vazifasida keladi. Eganing aniqlashtiruvchilari ot, sifat, son, olmosh, kiritma va yordamchi vazifasida keluvchi soʻzlar (undov, taqlid, modal) orqali ifodalangan aniqlovchi gap boʻlagidir. (Nurmonov, 1988, 110-bet) Kesim guruhida asosiy mazmuni fodalovchi unsur feʼl-kesim hamda ot-kesim hisoblanadi, shuningdek, turli soʻz turkumlari bilan ifodalangan toʻldiruvchi va hol kesimning aniqlashtiruvchilari boʻlib keladi. Mazkur modelda ega va kesim gapning bosh boʻlaklari, gapning semantik sintaktik yadrosi, aniqlovchi, toʻldiruvchi, hol esa ikkinchi darajali boʻlaklar deb nomlanadi.

**Aniqlovchi**— gapning ikkinchi muhim boʻlagi hisoblanadi. Jonli va jonsiz predmetga tobelanib, uning belgisini anglatadi. Aniqlovchining uch turi mavjud: 1) izohlovchi aniqlovchi 2) sifatlovchi aniqlovchi 3) qaratuvchi aniqlovchi).

**Toʻldiruvchi** — istalgan gapda ot yoki otlashgan soʻz bn ifodalanadigan, harakatning obʻyekti boʻlgan predmetni koʻrsatib, asosan kesimga bogʻlanadigan boʻlagi hisoblanadi. Boshqaruvchi soʻzga bogʻlanish xarakteriga koʻra 2 xil boʻladi; vositasiz va vositali toʻldiruvchi.

**Hol** -oʻzbek tilshunosligida gapning ikkinchi darajali boʻlaklaridan biri boʻlib, ish-harakatning belgi-sifatini, uning bajarilish usulini, shu bajarilish bilan bogʻliq boʻlgan oʻrin, payt, sabab, maqsad, shart-sharoit, miqdordaraja kabi xususiyatlarini koʻrsatadi. Hol ravish, feʼl, ot, sifat, son, olmosh kabi soʻz turkumlari bilan, shuningdek, taklidiy soʻzlar bilan ifodalanadi. Oʻzbek tilshunosligida hol gap boʻlagi 5 guruhga boʻlinadi: vaziyat(tarz)holi, payt holi, oʻrin holi, sabab holi, maqsad holi kabi.

Oʻzbek tilidagi istalgan matnni dasturlash tilida qayta ishlashga tayyorlash jarayoni bir nechta pagʻonalardan iborat boʻladi. Dastlab soʻzning qaysi turkumga mansub ekanligi morfologik analizator yordamida aniqlab olinadi.. Shundan soʻng berilgan istalgan matn uzunligiga qarab yuzaga chiqishi mumkin boʻlgan barcha holatlar aniqlanadi. Keyingi bosqichda grammatikasiga asoslangan oʻzbek tiliga ishlab chiqilgan muayyan qoidalarga koʻra yuqoridagi holatlar soni kamaytirib boriladi. Ishning soʻngida aniq natijaga ega boʻlgan bir necha holatlarga qoladi va shundan soʻng yuzaga kelgan holatlar foydalanuvchiga tanlovlar sifatida taqdim qilinadi. Shuni taʼkidlab ham oʻtish kerakki, qilinayotgan loyiha inversiyaga uchramagan hamda tarkibida 5 tagacha soʻz mavjud boʻlgan sodda gaplarning ehtimoliy kombinatsiyalarini taqdim qiladi. Aynan ushbu kombinatsiyalarni amalga oshirish uchun Python dasturlash tilining NLTK kutubxonasidan foydalanamiz. Pythonning NLTK kutubxonasida biz juda koʻplab jarayonlarni amalga oshirishimiz mumkin. Masalan, lugʻat va qoida kiritish orqali gaplarni sintaktik tahlil qilish ham

mumkin.. NLTK kutubxonasi gaplarni tahlil qilish uchun ,avvalo, morfologik va sintaktik teglar ishlab chiqilishi va gap bo‘laklarining kombinatsiyalari sonini aniqlab olish zarur bo‘ladi. Buning uchun avvalo, NLTK kutubxonasi Python dasturlash tiliga gaplardagi so‘zlarning qaysi turkumga tegishli ekanini va aynan shu bo‘lak gapda qaysi gap bo‘lagi vazifasini bajarayotganini tushuntirish uchun maxsus teglar ishlab chiqildi. So‘zlarni nutq qismlariga ajratish va ularni tegishli ravishda belgilash jarayoni nutq qismini teglash, POS-teglash yoki oddiygina teglash deb nomlanadi. Nutq qismlari so‘z turkumlari yoki leksik kategoriyalar deb ham ataladi. Muayyan vazifa uchun ishlatiladigan teglar to‘plami teglar to‘plami deb nomlanadi.

### Morfologik teglar:

№	Teg	So‘z turkumi
1	OT	Ot
2	SF	Sifat
3	SN	Son
4	OL	Olmosh
5	RV	Ravish
6	FL	Fe‘l

### Sintaktik teglar:

№	Teg	Gap bo‘lagi
1	G	Gap
2	EGA	Ega
3	KE	Kesim
4	AN	Aniqlovchi
5	TL	To‘ldiruvchi
6	HL	Hol
7	UND	Undalma
8	KRT	Kiritma

### 3.2-jadval. O‘zbek tili grammatikasi uchun yaratilgan teglar

Yuqorida yaratilgan teglar orqali loyiha hujjatlaridan misollar olib, ularni dasturlash tillariga o‘tkazish uchun har bir so‘zni teglanganiga misol keltiramiz:

**Morfologik teglarga misol:** Korxonalar va tashkilotlarning ro‘yxati doimo yangilanib boriladi.

Korxonalar - OT

Doimo - RV

Tashkilotlarning - OT

Yangilanib boriladi - FL

Ro‘yxati - OT

**Sintaktik teglarga misol:** Korxonalar va tashkilotlarning ro'yxati doimo yangilanib boriladi.

Korxonalar- AN

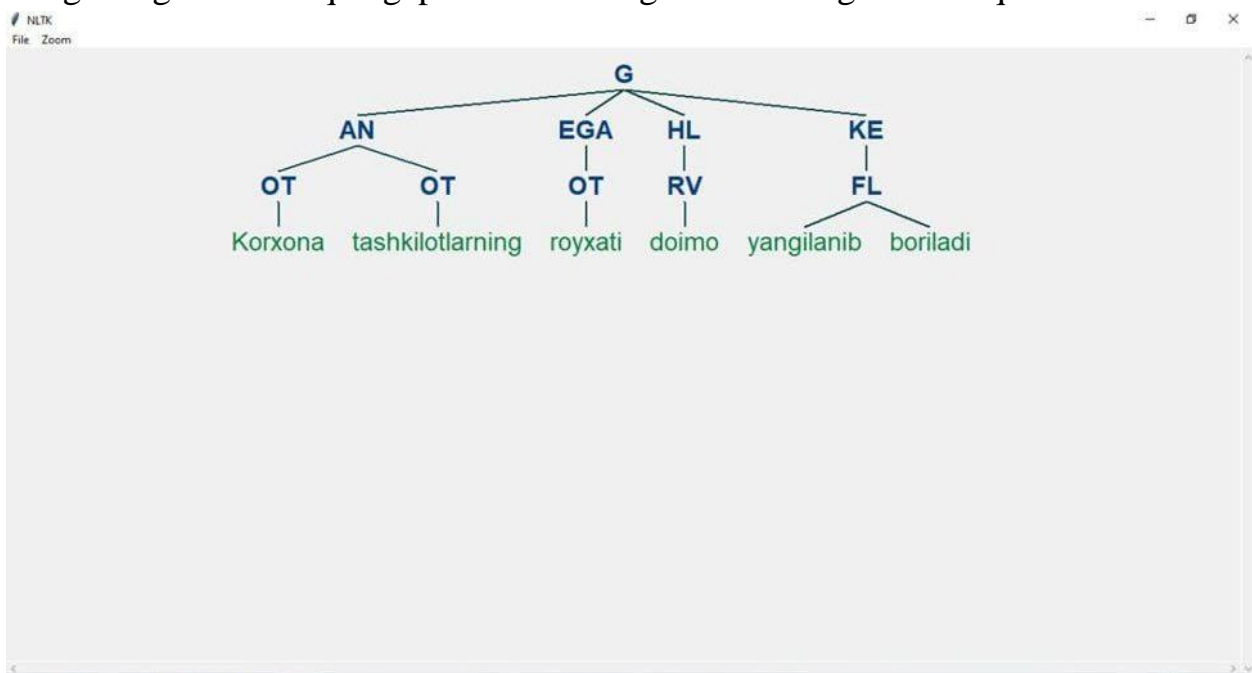
Doimo - HL

Tashkilotlarning-AN

Yangilanib boriladi-KE

Ro'yxati-E

Avvalo, yuqorida misolda berilgan gapning morfologik analizator lug'at qismi va sintaktik tahlil uchun grammatikaga asoslangan qoidalar tizimi ishlab chiqiladi. Qoidalar yaratib, ularni kiritilishining asosiy foydali joyi shundaki, ular taxmin qilinayotgan bo'lishi mumkin bo'lgan hodisalar sonini sezilarli darajada kamaytirishga xizmat qiladi. Qoidalar yaratilgan so'ng, tahlil daraxtida yuzaga keladigan bir gapning yaratgan teglarimiz orqali gap bo'laklarining ketma-ketligini hosil qilib olamiz.



Xulosa qilib shuni aytish mumkinki, ushbu maqolada tabiiy tillarning muhim bo'lagi bo'lgan KS-grammatika, sintaktik va morfologik teglar orqali sodda gaplarning tahlil daraxtini qanday qurishini ko'rib chiqdik:

1. Avvol, o'zbek tilining sintaksisi haqida umumiy ma'lumotlar berildi
2. O'zbek tilining gap bo'laklari va uning turlari tavsiflandi.
3. Sodda gaplarni sintaktik tahlil qilish uchun morfologik va sintaktik teglar yaratildi va jadvali tuzildi.
4. Sintaktik tahlil modellari va uning dasturiy ta'minotini yaratish uchun Python dasturlash va C# tilida sodda gaplardan iborat kichik dasturiy ta'minot yaratilgan va eng so'nggi bosqichimizda Tree.draw() funksiyasi yaratib olgan teglar bazamiz yordamida berilgan sodda gapimizning daraxti qurildi.

Ushbu mavzu bo'yicha ko'rib chiqilgan masalalar tabiiy tilni qayta ishlash jarayonida o'zbek tilining rivojlanishiga va takomillashishiga katta turki bo'ladi deb hisoblayman.

#### **FOYDALANILGAN ADABIYOTLAR RO'YXATI: (REFERENCES)**

1. Манушкин Евгений Сергеевич. (2012) Метод автоматического предсинтаксического анализа проектной документации с использованием к-грамматика. Москва. (24-25-betlar)
2. M. Sharipov, E. Kuriyozov, O. Yuldashev, and O. Sobirov, (2023). "UzbekTagger: The rule-based POS tagger for Uzbek language," arXiv preprint arXiv:2301.12711,.
3. Matlatipov, G., & Vetulani, Z. (2009, October). Representation of Uzbek Morphology in Prolog. In Aspects of Natural Language Processing (pp. 83-110).
4. Sadullayev, I. (2013). Tabiiy tillar sintaksisi uchun formalizmlar (O'zbek tili misolida) magistrlik dissertatsiyasi. Buxoro. (46-47-betlar)
5. A. Nurmonov. (1988) Tilni sistemali o'rganish va sintaksisning ayrim munozarali masalalari. O'zbek tili va adabiyoti. (15-bet)
6. Maksud Sharipov, Jamolbek Mattiev, Jasur Sobirov, Rustam Baltayev. (2022) Creating a Morphological and Syntactic Tagged Corpus for the Uzbek Language, CEUR Workshop Proceedings (14-bet)