

O'ZBEK TILIDAGI SODDA GAPLAR UCHUN SINTAKTIK PARSING MODELI

Raxmonova Mohira Abdusattor qizi
ToshDO'TAU Kompyuter lingvistikasi
mutaxassisligi 2-kurs magistranti
mokhirata@gmail.com

Ilmiy rahbar: f.f.d. (DSc) Sh.Hamroyeva

Annotatsiya. Ushbu maqolada sintaktik analizatorning tahlil qilish algoritmi va tahlil qilish bosqichlari, o'zbek tilidagi sodda gaplarning o'zbekcha va inglizcha modellari ishlab chiqilgan. Buning uchun, avvalo, gaplarning sintaktik qoidalari formal grammatikadan foydalangan holda tavsiflanadi. Rasmiy grammatika sifatida Xomskiyning Context free grammatikasidan foydalanilgan.

Kalit so'zlar: *sintaktik analizator algoritmi, sodda gap modellari, Context free grammatikasi, POS tagging.*

Аннотация. В данной статье разработаны алгоритм и этапы анализа синтаксического анализатора, узбекская и английская модели простых узбекских предложений. Для этого, прежде всего, с помощью формальной грамматики описываются синтаксические правила предложений. Context free грамматика Хомского использовалась в качестве официальной грамматики.

Ключевые слова: *алгоритм синтаксического анализатора, модели простых предложений, Контекст-фрии грамматика, POS tagging.*

Abstract. In this article, an algorithm and stages of analysis of the Syntactic parser, Uzbek and English models of simple Uzbek sentences are developed. To do this, first of all, with the help of formal grammar, the syntactic rules of sentences are described. The Context free grammar of Chomsky was used as the official grammar.

Keywords: *Syntactic parser algorithm, simple sentence models, context-free grammar, POS tagging.*

Formal tillar va formal grammatika nazariyasi XX asrning 50-yillarida N.Xomskiyning generativ nazariyasidan kelib chiqdi. Ushbu nazariya sintaktik tuzilmalarini tahlil qilish uchun asos bo'ldi [3]. Avtomatik sintaktik tahlil (parsing) kompyuter lingvistikasi nazariyasining eng ko'p o'rganilayotgan sohasidir. Sintaktik analizator ma'lumotlar bazasi interfeyslarini loyihalashtirish, sun'iy intellekt, matnlarga avtomatik ishlov berish, avtomatlashtirilgan ma'lumot-qidiruv tizimi sohasi, mashina tarjimasida keng qo'llanilmoqda. Ta'kidlash kerakki, avtomatik sintaktik tahlil (ST) usullari ishi tabiiy tilning formal sintaktik nazariyalaridan kelib chiqadi [1; 78].

Sintaksis deganda tilshunoslikda asosiy birligi gap va eng kichik birligi so'z birikmalari bo'lgan sath nazarda tutilgan. Unda gap tarzida muayyan hukmni ifoda qiladigan grammatik izchil bog'langan so'zlar zanjiri tushuniladi. Grammatik izchil

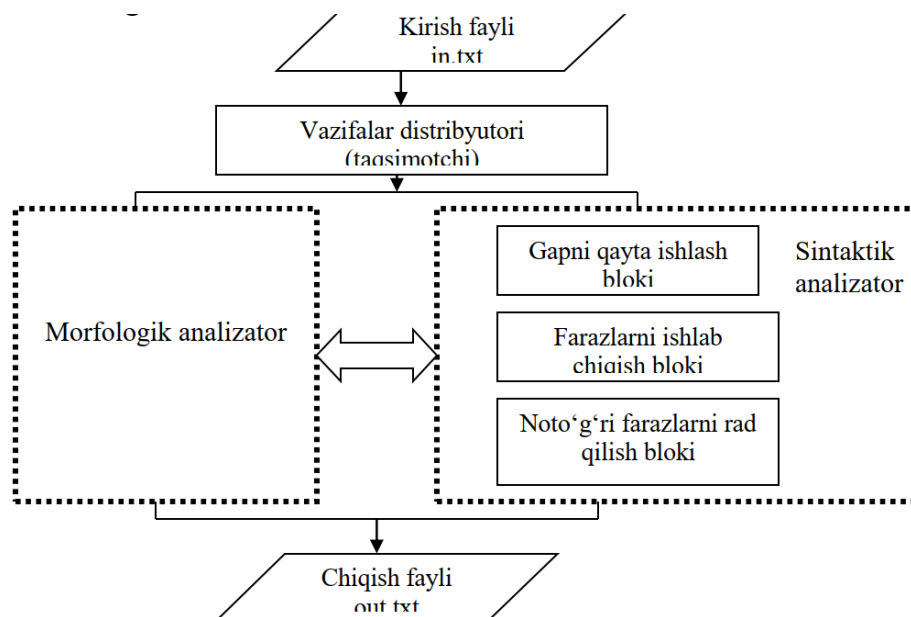


zanjir – so‘zshakllarning o‘zaro ma’lum bir grammatik munosabatlarda bog‘lanishidir. So‘zshakl yoki grammatik so‘z leksemaning o‘ziga biriktirishi mumkin bo‘lgan grammatik shaklda kelishidir. Shunday qilib, gapning sintaktik tuzilishi so‘z shakllarning o‘zaro sintaktik bog‘langan zanjiridan tashkil topadi [1; 78].

M.Abjalovanning ko‘rsatishicha, sintaktik tahlil moduli (STM) sintaktik tahlil qilish dasturi yoki lingvistik dasturning sintaktik tahlilni amalga oshirishga mo‘ljallangan qismi. Bugungi kunda avtomatik STMlarini yaratish kompyuter lingvistikasining dolzarb masalalaridan hisoblanib, buning yechimi turli amaliy maqsadlarni yuzaga keltirish uchun til sistemasini formallashtirishda yuqori sifatli natijalar imkonini beradi: til va nutq birliklarini avtomatik tekshirish sistemasini yaratish, Internetning qidiruv sistemasini mukammallashtirish shular jumlasidandir. Sintaktik tahlildan maqsad sintaktik strukturalar va sintaktik bog‘lanish me‘yorlarini kuzatishdir, biroq barcha tillarda ham STM yaratish murakkab ish hisoblanadi, chunki amaliy tilshunoslikda yetarlicha nazariy va amaliy tadqiqotlar mavjud emas; til va nutq birliklarining sintaktik bog‘lanishi individual xilma-xillikka ega [1; 82].

Har bir tilning sintaktik analizatori (parser) o‘sha til tabiati va qonuniyatlariga binoan yaratiladi. Parser ishi so‘zlar kategoriyasi (turkumlar) ma’lumotlari bazasiga asoslanadi va asosiy vazifasi ayrim tillarda jumlada gap bo‘laklarini aniqlash (mas., rus tilida) hisoblansa, muayyan tillarda, masalan, ingliz tilida N.Xomskiy nazariyasiga asosan otli guruh va fe‘lli guruh tarzida generativ tahlil qilinadi [1; 88-89].

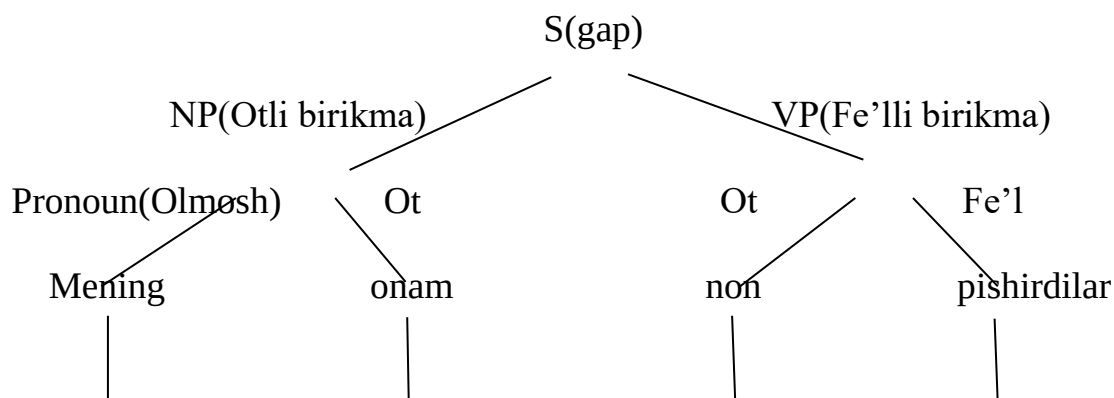
Quyida A.Leonteva va I.Kagirovlar tomonidan ishlab chiqilgan analizator dasturi blok-sxemasi berilgan. Bunda gaplarni sintaktik tahlil qilish bosqichlari keltirilgan.



1-chizma. A.Leonteva va I.Kagirovlar tomonidan analizator dasturi blok-sxemasi[6]

Kirish faylni oddiy gaplar tashkil etadi. Vazifalarni taqsimlovchi ma'lumotlarni tahlil qiluvchiga uzatadi. Ishlov berish blokida gaplar so'zma-so'z qayta ishlanadi. Asl so'z shakli morfologik tahlil blokiga topshiriladi, unda o'zaklarning barcha mumkin bo'lgan variantlari va unga mos keladigan grammatik ko'rsatkichlar tanlanadi. Har bir gap so'zma-so'z o'qiladi, shundan so'ng so'z shakli morfoanalizatorga kiritiladi. Natijada barcha mumkin bo'lgan o'zaklar va tegishli grammatik ko'rsatkichlar aniqlanadi. Agar berilgan so'z shakliga faqat bitta asos mos kelsa, u gipotezalarni qurish tartibiga kiradi. Gap qismiga va grammatik ko'rsatkichlarga ko'ra tegishli sintaktik guruh ajratiladi. Dastlabki ishlovdan so'ng so'z shakli gipoteza yaratish birligiga kiradi. Ushbu blok asosiy hisoblanadi. Uning kiritilishi so'z shaklidir. Bu gapdagi birinchi so'z bo'lsa, nutqqa mos ravishda sintaktik guruh aniqlanadi. Shunday qilib, morfologik va sintaktik noaniqlik tufayli bir gap uchun bir nechta farazlar tuzilishi mumkin. Bundan tashqari, bu farazlar noto'g'ri gipotezalarni tekshirish blokiga kiradi. Ushbu blokda tekshirishning ikki darajasi mavjud. Birinchi bosqichda sintaktik guruhlarning bir gap doirasidagi kelishiklari tekshiriladi. Chiqarish fayli gaplar ro'yxati bo'lib, ularning har biri sintaktik guruhlarga bo'linadi. Agar gapda lug'atda mavjud bo'lmagan so'z bo'lsa, u to'g'ridan-to'g'ri ko'rsatiladi [3; 399].

Sintaktik tahlilda jumla elementlarining sintaktik munosabatlari aniq belgilab qo'yilgan sintaktik daraxt, ya'ni tobelik daraxtiga asoslaniladi. Masalan: *Mening onam non pishirdilar* jumlasida quyidagicha tobeli qismlarga ajratilib, turkumlari bo'yicha tahlil qilinadi:



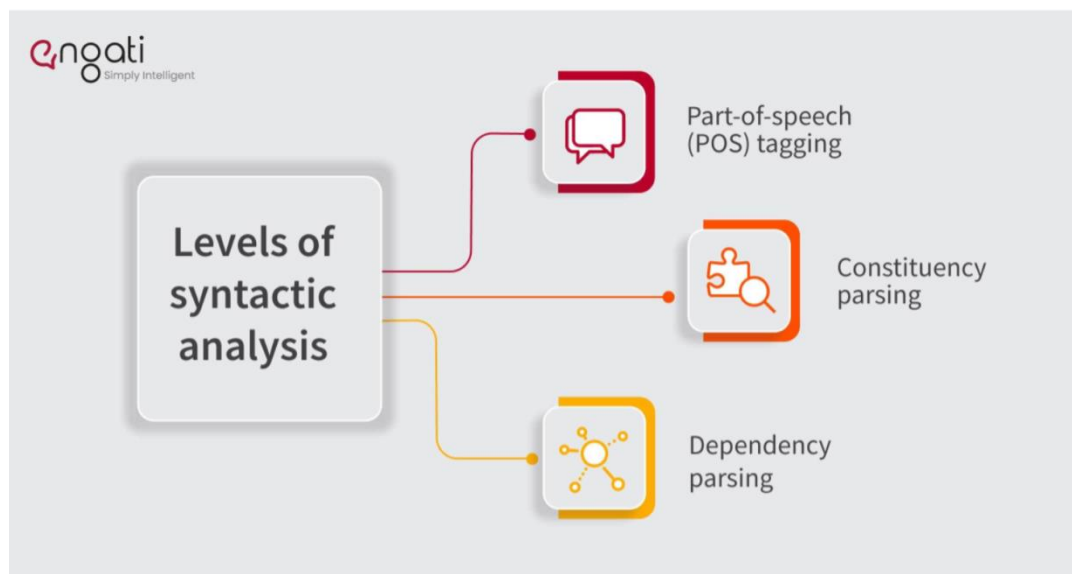
Bevosita ishtirokchilar (BI) metodiga ko‘ra tayanch nuqta konstruksiya (qurilma) deb ataladi. Konstruksiya tarkibidan bevosita ishtirokchilar (BeI) va bilvosita ishtirokchilar (BiI) ajratiladi [1; 24-25].

O‘zbek tili sintaktik analizatorining sintaktik tahlil algoritmi

An’anaviy grammatik vositalar yordamida gapni sintaktik tahlil qilishda quyidagi qadamlarni bajarish mumkin:

- 1) gap bo‘laklari chegaralari va so‘z chegaralarini aniqlash
- 2) bo‘laklarini aniqlash;
- 3) tarkibiy qismlarni aniqlash;
- 4) tarkibiy qismlar uchun sintaktik kategoriyalarni aniqlash;
- 5) tarkibiy qismlarning grammatik vazifalarini aniqlash;
- 6) sintaktik strukturani chizish;
- 7) noto‘g‘ri tuzilgan gaplarni to‘g‘ri tartibga solish.

Sintaktik tahlilning maqsadi. Uning maqsadi kiritilgan matnining tuzilishini eng kichik asosiy bo‘laklardan tortib gaplarga tushunish va undan mantiqiy ma’no chiqarishdir. Sintaktik tahlil tabiiy tilni qayta ishlashning (Natural Language Processing) juda muhim jihati hisoblanadi, chunki u har qanday jumlaning grammatik ma’nosini aniqlashga yordam beradi [4].



3-chizma. Sintaktik tahlil bosqichlari[2]

Sintaktik tahlil bosqichlari Part-of-speech tagging, Constituency parsing va Dependency parsing lardan iborat.

Gap bo‘laklarini (Part-of-speech) teglash. Bu sintaktik tahlilning birinchi bosqichidir. Gap bo‘lagini teglash sintaktik tahlilning muhim qismi bo‘lib, gapdagi so‘z (fe‘l, ot, sifat va boshqalar) va qo‘shimchalar sifatida belgilashdan iborat. Gap bo‘laklarini teglash gapning ma‘nosini tushunishimizga yordam beradi. Boshqa barcha tahlil usullari gap bo‘laklari teglaridan foydalanadi.

Guruhlarga ajratib tahlil qilish (Constituency Parsing). Guruhlarga ajratib tahlil qilish so‘zlarning gapdagi grammatik maqomiga ko‘ra, gapdan guruhlarga, ya‘ni kichik iboralarga ajratishni o‘z ichiga oladi. Otli birikmalar (Noun Phrases) va fe‘lli birikmalar (Verb Phrases) eng keng tarqalgan birikmalar hisoblanadi.

Tobe bog‘lanish (Dependency Parsing). Tobe bog‘lanish erkin so‘z tartibidagi tillarda keng qo‘llaniladi. Tobelikni tahlil qilishda so‘zlarning o‘rtasida bog‘lanish hosil bo‘ladi. Ikki so‘z o‘rtasida bog‘liqlik mavjud bo‘lsa, bir so‘z hokim, ikkinchisi esa tobe bo‘lak bo‘ladi [2].

O‘zbek tilida sodda gap

Sodda gap ma‘lum fikr, maqsadni ifodalovchi, grammatik va intonatsion tomondan shakllangan sintaktik birlikdir. Sodda gap tuzilishiga ko‘ra yig‘iq yoki yoyiq bo‘lishi mumkin. Bosh bo‘laklardan tarkib topgan gaplar – yig‘iq gap, agar bosh bo‘laklardan tashqari ikkinchi darajali bo‘laklar ham ishtirok etsa, yoyiq gap sanaladi. O‘zbek tilida so‘zlar tartibi asosan erkin bo‘lsa ham, sodda gapda har bir gap bo‘lagining odatda qo‘llanadigan o‘z o‘rni bor. Bu gapning strukturasi, gap bo‘lagining sintaktik ifodalanish usuli, u bilan bevosita bog‘liq bo‘lgan boshqa so‘zlarning o‘rni kabi hodisalar bilan belgilanadi. Shu asosda so‘zlarning to‘g‘ri (odatdagi) tartibi va teskari (odatdagidan chekinish) tartibi farq qilinadi [5; 179].

O'zbek tilida sodda gap modellari



Ushbu jadvalda 32 ta sodda gaplarning modellari keltirilgan. Sodda gap modellari o'zbek va ingliz tilida berilgan. Bundan tashqari, har bir gap modeli uchun Context free grammatikasi ishlab chiqilgan.

Context free Chomskiy grammatikasi (CFG) NLP da tarkibiy tuzilmalarni modellashtirish uchun eng keng tarqalgan rasmiy tizimdir. CFG bir qator qoidalar iboratdir. CFG G to'rtta parametr bilan belgilanadi. Context free grammatika - bu ma'lum bir rasmiy tilda barcha mumkin bo'lgan gaplarni yaratish uchun foydalaniladigan rasmiy grammatika.

Context free grammatikasida G to'rtta belgidan foydalaniladi: $G = (V, T, P, S)$

G – grammatika

T – terminal belgilarning cheklangan to'plamini

V – terminal bo'lmagan belgilarning cheklangan to'plami

P – ishlab chiqilgan qoidalari to'plamini

S – boshlanish belgisi.

E – ega, K – kesim, A – aniqllovchi, H – hol, T – to'ldiruvchi, S – subject, O – object, Adj – adjective, Adv – adverb, V – verb

№	Sodda gaplar	O'zbek tilidagi model	Ingliz tilidagi model	Context Free Grammar
1	Poyezd keldi	E+K	S+V	$S \rightarrow NP VP$ $NP \rightarrow S$ $VP \rightarrow V$ $S \rightarrow Poyezd$ $V \rightarrow keldi$
2	Dalaga chiqmaysizmi?	H+K	Adv+V	$S \rightarrow VP$ $VP \rightarrow Adv V$ $Adv \rightarrow dalaga$ $V \rightarrow chiqmaysizmi$



3	Kitobni o'qidi	T +K	O+V	S->VP VP->O V O->kitobni V->o'qidi
4	Anna kecha omadli edi	E+H+K	S+Adv+V	S->NP VP NP->S Adv S VP->Adv V V S->Anna Adv->kecha V->omadli edi
5	Men taqdimotga kechikdim	E+H+K	S+Adv+V	S->NP VP NP->S VP->Adv V S->Men Adv->taqdimotga V->kechikdim
6	Sabrina kecha to'pni dribling qildi	E+H+T+K	S+Adv+O+V	S->NP VP NP->S Adv S VP->Adv O V O V S- >Sabrina Adv->kecha O- >to'pni V->dribling qildi
7	Ali kimyo kitobini o'qiyapti	E+A+T+K	S+Adj+Adv+V	S->NP VP NP->S VP->Adj O V S->Ali Adj->kimyo O- >kitobini V->o'qiyapti
8	Men uchun quyosh chiqishi o'ziga xosdir	T+E+K	O+S+V	S->NP VP NP->O S VP-> S V O V S->quyosh chiqishi O->men uchun V- >o'ziga xosdir
9	Ali bu yomon qo'shiqni kuyladi	E+A+A+T+K	S+Adj+Adj+O+V	S->NP VP NP->S VP->Adj Adj O V S->Ali Adj->bu yomon O->qo'shiqni V- >kuyladi
10	Ali qo'shiqni kuyladi	E+T+K	S+O+V	S->NP VP NP->S VP->O V S->Ali O->qo'shiqni V- >kuyladi
11	Talabalar iborani bir oyda bitirdilar	E+T+H+K	S+O+Adv+V	S->NP VP NP->S VP->O Adv V S->Talabalar O- >iborani Adv->bir oyda V- >bitirdilar
12	Import qilingan mashinalar qimmat	A+E+K	Adj+S+V	S->NP VP NP->Adj S VP- >V Adj->Import qilingan S- >mashinalar V->qimmat
13	Bu muzqaymoq brendi ajoyib	A+E+K	Adj+S+V	S->NP VP NP->Adj S VP- >V Adj->bu S->muzqaymoq brendi V->ajoyib



14	Oq bulutlar bilan qoplangan	A+T+K	Adj+O+V	S->VP VP->Adj O V Adj->Oq O->bulutlar bilan V->qoplangan
15	Murat Orhun ertaga mashg'ulot o'tkazadi	E+H+T+K	S+Adv+O+V	S->NP VP NP->S VP->Adv O V S->Murat Orhun Adv->ertaga O->mashg'ulot V->o'tkazadi
16	Bu gullar ko'p suvga muhtoj emas	A+E+A+T+K	Adj+S+Adj+O+V	S->NP VP NP->Adj S VP->Adj O V Adj->bu ko'p S->gullar O->suvga V->muhtoj emas
17	Bir o'rdak daryoni kesib o'tdi.	A+E+T+K	Adj+S+O+V	S->NP VP NP-> Adj S VP->O V Adj-> bir S->o'rdak O->daryoni V->kesib o'tdi
18	Ali kun bo'yi qo'shiq kuyladi	E+H+T+K	S+Adv+O+V	S->NP VP NP->S VP->Adv O V S->Ali Adv->kun bo'yi O->qo'shiq V->kuyladi
19	Karla baxtsiz xodisaga uchradi	E+A+T+K	S+Adj+O+V	S->NP VP NP->S VP->Adj O V S->Karla Adj->baxtsiz O->xodisaga V->uchradi
20	Men kinoga dugonam bilan bordim	E+H+T+K	A+Adv+O+V	S->NP VP NP->S VP->Adv O V S->men Adv->kinoga O->dugonam bilan V->bordim
21	Gulnora kecha ishga kelmadi	E+H+H+K	S+Adv+Adv+V	S->NP VP NP->S VP->Adv Adv V S->Gulnora Adv->kecha ishga V->kelmadi
22	Nilufar onasi uchun chiroyli gullar sotib oldi	E+T+A+T+K	S+O+Adj+O+V	S->NP VP NP->S O VP->Adj O V S->Nilufar O->onasi uchun gullar Adj->chiroyli V->sotib oldi
23	Ushbu ko'cha doim gavjum bo'ladi	A+E+H+K	Adj+S+Adv+V	S->NP VP NP->Adj S VP->Adv V Adj->ushbu S->gullar Adv->doim V->gavjum bo'ladi
24	Men kecha universitetga Hamid bilan keldim	E+H+H+T+K	S+Adv+Adv+O+V	S->NP VP NP->S VP->Adv Adv O V S->men Adv->kecha universitetga O->Hamid bilan V->keldim



25	Men kecha yaqin do'stim bilan universitetga keldim	E+H+A+T+H +K	S+Adv+Adj+O+Adv +V	S->NP VP NP->S Adv VP- >Adj O Adv V S->men Adv- > kecha universitetga Adj- >yaqin O->do'stim bilan V- >keldim
26	Biz Hamidning uyida imtihonlarga tayyorlandik	E+A+H+T+K	S+Adj+Adv+O+V	S->NP VP NP->S Adj Adv VP->O V S->biz Adj- >Hamidning Adv->uyida O- >imtihonlarga V- >tayyorlandik
27	Men sendan kitobimni kecha olmoqchi edim.	E+T+T+H+K	S+O+O+Adv+V	S->NP VP NP-> S O VP->O Adv V S->men O->sendan kitobimni Adv->kecha V- >olmoqchi edim
28	Men hech qachon hech kimdan hech narsa kutmayman	E+H+T+T+K	S+Adv+O+O+V	S->NP VP NP-> S Adv VP- >O O V S->men Adv->hech qachon O->hech kimdan hech narsa V->kutmaymanm
29	Men kecha universitetga Hamidning oyisi bilan keldim	E+H+H+A+T +K	S+Adv+Adv+Adj+O +V	S->NP VP NP->S Adv Adv VP->Adj O V S->men Adv- >kecha universitetga Adj- >Hamidning O->oyisi bilan V->bordimm
30	Biz do'stimning akasi bilan universitetga birga vaqtliroq keldik	E+A+T+H+H +H+K	S+Adj+O+Adv+Adv +Adv+V	S->NP VP NP->S Adj O VP->Adv Adv Adv V S->biz Adj->do'stimning O->akasi bilan Adv- >universitetga birga vaqtliroq V->keldik
31	Men senga bu buyumga qo'llaring bilan tegmasligingni bir necha bor aytgandim	E+T+A+T+T+ T+H+K	S+O+Adj+O+O+O+ Adv+V	S->NP VP NP->S O Adj O VP->O O Adv V S->Men O- >senga buyumga qo'llaring bilan Adj->bu Adv-> bir necha bor V->aytganman

Xulosa sifatida aytish joizki, o'zbek tili uchun yaratilgan sintaktik analizator sintaktik tahlil va parsing algoritmi uchun, tilshunoslarning ilmiy ishlari hamda boshqa analizatorlar yaratishda asos bo'lib xizmat qiladi. Sodda gaplar uchun tuzilgan o'zbekcha va inglizcha modellar asosida sintaktik analizator yaratish uchun algoritmlar ishlab chiqiladi.

Foydalanilgan adabiyotlar:

1. B.Elov, Sh.Hamroyeva, Z.Xusainova. NLP (tabiiy tilga ishlov berish)ning vazifalari va zamonaviy yondashuvlar. TerDU, filologik tadqiqotlar: til, adabiyot, ta'lim. 2022, 5-6.
2. O.Abdullayeva. Til korpuslarida lingvistik annotatsiya va uning prinsiplari. O'zbek tilining milliy korpusi: muammo va vazifalar mavzusidagi xalqaro ilmiy-amaliy anjumani materiallari. – Toshkent, 31-may 2022-yil. – B. 130-136.
3. O.Abdullayeva. O'zbek tili korpusida matnlarni sintaktik annotatsiyalash masalasi. // O'zbek amaliy filologiyasi istiqbollari respublika ilmiy-amaliy anjumani materiallari. – Toshkent, 26-oktabr 2022-yil. – B. 122-126.
4. <https://www.engati.com/glossary/syntactic-analysis>
5. https://www.hse.ru/data/2017/07/24//КЛ_задачи_подходы_ресурсы.
6. [https://www.tokenex.com/blog/ab-what-is-nlp-natural-language-processing-tokenization/#:~:text=Tokenization%20is%20used%20in%20natural,into%20understandable%20parts%20\(words\)](https://www.tokenex.com/blog/ab-what-is-nlp-natural-language-processing-tokenization/#:~:text=Tokenization%20is%20used%20in%20natural,into%20understandable%20parts%20(words)).
7. O'zbek tili grammatikasi. II t. Sintaksis. Mas'ul muharrirlar G'.A.Abdurahmonov, Sh.Shoabdurahmonov, A.Hojiyev. – Toshkent. 1976. – B. 179.
8. Леонтьева А., Кагиров И. Автоматический синтаксический анализ русских текстов // Труды 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2008, Дубна, Россия, 2008. – С. 399.