

NLP'DA SINTAKTIK TAHLILNING O'RNI

Abdullayeva Oqila Xolmo'minovna,
Doktorant DSc
abdullayevaqila@gmail.com
ToshDO'TAU

Annotatsiya. Tabiiy tilni qayta ishlash (Natural Language Processing – NLP) sun'iy intellektning eng muhim yo'nalishlaridan biri bo'lib, inson tilini kompyuter yordamida qayta ishlash, tushunish va generatsiya qilish masalalarini o'rganadi. Zamonaviy NLP tizimlari avtomatik tarjima, matn tahlili, savol-javob tizimlari, chatbotlar va axborot qidiruv tizimlarida keng qo'llanmoqda. Ushbu tizimlarning samaradorligi esa til birliklari o'rtasidagi grammatik va semantik munosabatlarni to'g'ri aniqlashga bog'liq. Mazkur maqolada o'zbek tili matnlarini avtomatik sintaktik tahlil qilishning va sintaktik analizatorlarni yaratishning NLPda tutgan o'rni tahlil qilinadi. Sintaktik analizatorlarning samaradorligi va aniqligi bevosita sifatli korpus, treebank va lingvistik resurslarning mavjudligiga bog'liq ekanligi asoslab beriladi. Tadqiqotda dependency parsing, Universal Dependencies (UD), morfologik teglash, sintaktik annotatsiya, treebank yaratish metodologiyasi hamda o'zbek tilining agglyutinativ tabiati bilan bog'liq muammolar atroflicha yoritilgan. Shuningdek, parserlarni yaratishda reprezentativ korpuslarning roli, annotatsiya standartlari, neyron tarmoqlar va transformer modellaridan foydalanish istiqbollari muhokama qilinadi. Maqola natijalari o'zbek tilida zamonaviy sintaktik analizatorlarni yaratish uchun milliy korpus va lingvistik ta'minot tizimini shakllantirish zarurligini ko'rsatadi.

Kalit so'zlar: *sintaktik tahlil, parsing, dependency parsing, analizator.*

Abstract. Natural Language Processing (NLP) is one of the most important areas of artificial intelligence, studying the issues of processing, understanding, and generating human language using a computer. Modern NLP systems are widely used in automatic translation, voice assistants, chatbots, text analysis, Q&A systems, and



information retrieval systems. The effectiveness of these systems depends on the correct determination of grammatical and semantic relations between language units. This article analyzes the role of automatic syntactic analysis of Uzbek texts and the creation of syntactic analyzers in NLP. It is substantiated that the efficiency and accuracy of syntactic analyzers directly depend on the availability of a high-quality corpus, tree bank, and linguistic resources. The study covers issues related to dependency parsing, Universal Dependencies (UD), morphological tagging, syntactic annotation, the methodology for creating a tree bank, and the agglutinative nature of the Uzbek language. The role of representative corpora in parser creation, annotation standards, and the prospects for using neural networks and transformer models are also discussed. The results of the article show the need to form a national corpus and a linguistic support system for creating modern syntactic analyzers in the Uzbek language.

Keywords: *syntax parsing, parsing, dependency parsing, parser, syntactic analyzer.*

Sintaktik tahlil (syntactic parsing) NLP tizimlarining markaziy komponentlaridan biri hisoblanadi. Chunki matndagi so‘zlarning o‘zaro bog‘lanishini aniqlamasdan turib matnning mazmunini to‘liq tushunish mumkin emas. Sintaktik tahlil yordamida gapning grammatik strukturasi, ega-kesim munosabati, tobe-hokim bog‘lanishlari va boshqa sintaktik aloqalar aniqlanadi. Sintaktik tahlilning NLPdagi o‘rni shundaki, u tilni qayta ishlashning quyi sathlari hisoblangan *tokenizatsiya, lemmatizatsiya, morfologik belgilash* bilan yuqori sathlari *semantik rolni aniqlash, savol-javob tizimlari, mashina tarjimasi, diskurs tahlili* o‘rtasida ko‘prik vazifasini bajaradi. Morfologik tahlil “*qaysi birlik qanday grammatik xususiyatga ega?*” degan savolga javob bersa, sintaktik tahlil “*ushbu birliklar gap ichida qanday munosabatga kirishgan?*” degan savolni hal etadi. Masalan, otning kelishik shakli, fe’ning shaxs-son ko‘rsatkichi yoki yordamchi



unsur mavjudligi morfologik darajada aniqlansa, aynan sintaktik bosqichda “kim?”, “nimani?”, “qayerda?”, “qaysi birlik qaysi birlikka hokim yoki tobe?” kabi munosabatlar tizimlashtiriladi. Shu sababli sintaktik tahlilsiz matnni chuqur tushunishga qaratilgan ko‘plab NLP vazifalari faqat yuzaki belgilar bilan cheklanib qoladi. UD’ning asosiy tamoyillari ham aynan grammatik bog‘lanishlarni universal abstraksiya sifatida olish, parsing uchun qulaylik yaratish vazifalarini qo‘llab-quvvatlashga qaratilgan.

Sintaktik tahlilning NLPdagi rolini to‘liq tushunish uchun grammatik munosabatlarning qaysi darajada abstraksiya qilinishi muhimligini ko‘rish zarur. UD doirasida grammatik relatsiyalar tilning yuzaki belgilaridan yuqoriroq, ammo to‘liq semantik formalizmdan pastroq oraliq sathda joylashadi. Bunday yondashuv, bir tomondan, tillararo qiyosni osonlashtiradi; ikkinchi tomondan esa relatsiyalarni kompyuter uchun amaliy jihatdan soddalashtiradi. de Marneffe va boshqalar ta’kidlaganidek, UD’da “*content words*” orasidagi munosabatlarni markazga chiqarish, funksional so‘zlarni esa ko‘pincha tobe unsur sifatida ko‘rish munosabatlarni ajratib olish jarayonini yengillashtiradi va predikat-argument strukturasi aniqroq namoyon etadi [1]. Shuning uchun zamonaviy parsingda sintaksisning vazifasi shunchaki “*grammatik daraxt*” qurish emas, balki NLP uchun eng foydali strukturani ajratib olishdan iborat.

Bu jihat, ayniqsa, analitik va agglyutinativ tillarda turlicha ahamiyat kasb etadi. Analitik tillarda funksional so‘zlar va so‘z tartibi kuchli signal vazifasini bajarsa, agglyutinativ tillarda affiksatsiya, kelishik, egalik va kesimlik shakllari parser qarorlarini sezilarli darajada boshqaradi. Shuning uchun sintaktik tahlil ko‘pincha “*sof sintaksis*” emas, balki morfosintaktik tahlil sifatida ishlaydi. CoNLL-2018 metrikalarida LAS bilan bir qatorda MLAS va BLEXning joriy etilgani ham parser sifatini morfologiya va lemmatizatsiyadan uzib bo‘lmasligini ko‘rsatadi [2].

Kompyuter lingvistikasida sintaktik tahlilning real ilmiy-amaliy bazasi, avvalo, *treebanklar, parserlar va tilning yordamchi resurslarining mavjudligi bilan* belgilanadi. Treebank annotatsiyalangan korpus sifatida sintaktik modelni o‘qitish, sinash va baholash uchun “*oltin standart*” manba vazifasini bajaradi; parser esa ana shu strukturani avtomatik tiklaydigan hisoblash mexanizmidir. Morfologik analizator, lemmatizator, POS-tagger, transliteratsiya vositalari va oldindan o‘qitilgan til modellari esa parserning kirish qatlamini va umumiy aniqligini belgilaydi. Bugungi kunda bunday infratuzilmaning eng muhim umumlashgan platformasi Universal Dependencies (UD) bo‘lib, u 150 dan ortiq til va 200 dan ortiq treebankni qamrab oladi, shu bilan birga, UD asosidagi parser ekotizimlari CoNLL-U formatini de-fakto xalqaro standart darajasiga olib chiqqan.

Quyidagi jadval parserlar evolyutsiyasining NLP nuqtayi nazaridan umumiy manzarasini ko‘rsatadi. Jadval MaltParser, MST parsing, early neural, BiLSTM/biaffine, multilingual pretraining va LLM-based parsing bo‘yicha asosiy ishlar asosida umumlashtirildi.

1-jadval. Sintaktik tahlil texnologiyalarining evolyutsiyasi

Bosqich	Vakil yondashuv	Asosiy g‘oya	Kuchli tomoni	Cheklovi
Qoidaviy bosqich	grammar-driven parserlar	Formal grammatika qoidalariga tayanuvchi tahlil	Interpretatsiyasi aniq	Keng ko‘lamli qoida bazasi talab qiladi
Statistik data-driven bosqich	MaltParser, MSTParser	Treebankdan o‘rganish	Tilga moslashuvi osonlashdi	Feature engineering’ga kuchli bog‘liqlik
Erta neyron bosqich	Chen & Manning	Neyron klassifikatsiya bilan transition qarorlarini tanlash	Tezlik va aniqlik yaxshilandi	Kontekst ifodasi hali cheklangan
Chuqur neyron bosqich	BiLSTM, biaffine	Kontekstual featurelarni	Yuqori aniqlik, kamroq qo‘l mehnati	Ko‘proq ma’lumot va



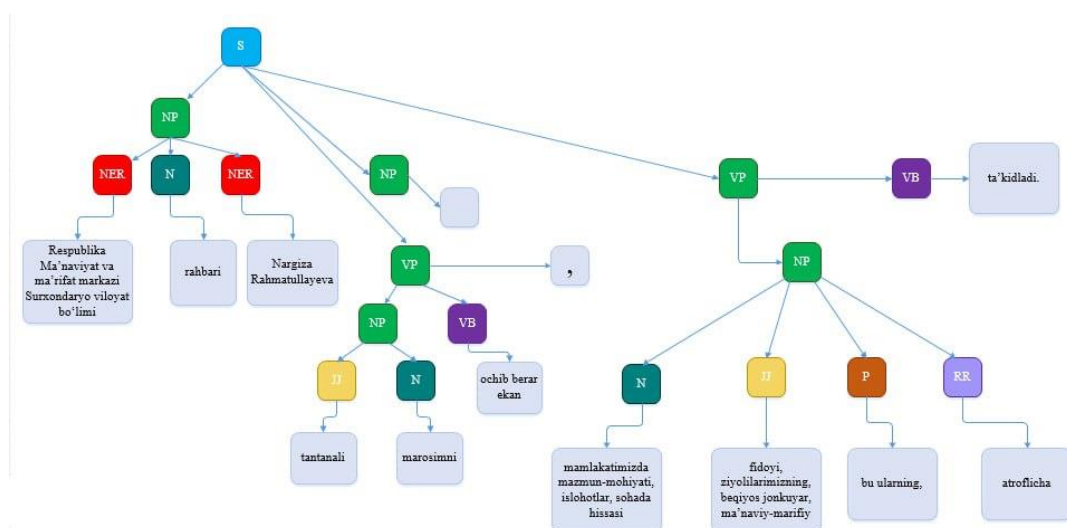
		avtomatik chiqarish		resurs talab qiladi
Multilingual/pretrained bosqich	UDPipe 2, UDify	Multi-task, cross-lingual, pretrained encoder	Kam resursli tillarda foydali	Transfer sifati tillar orasidagi farqqa bog'liq
LLM bosqichi	step-by- step tabular parsing	Tabiiy til ko'rsatmalari bilan strukturali chiqish hosil qilish	Yangi moslashuvchanlik va format birxilligi	Barqarorlik, validlik va nazorat masalalari saqlanadi

Bu jadvaldan anglashiladiki, sintaktik tahlil texnologiyasi o'zgargan sari uning NLPdagi vazifasi yo'qolmagan, aksincha kuchaygan. O'zgarish parserning “bor-yo'qligi”da emas, balki uning qanday qurilishi, qaysi resurslarga tayangani, qanday downstream tizimlarga ulanayotgani va qaysi universal standartda ifodalanganida namoyon bo'ladi.

Sintaktik parser sifati, ayniqsa, agglyutinativ tillarda, faqat treebankka emas, balki parserning old qatlamida ishlaydigan morfologik, leksik va neyron resurslarga ham bog'liq bo'ladi. So'nggi yillarda neyron modellar bu bazani yanada boyitdi. Turkiy tillar ichida turk tili resurslar jihatidan eng rivojlangan tayanch nuqta bo'lib turibdi. 2023-yildagi tanqidiy tahlilda turk tili turkiy tillar ichida eng yirik va eng yaxshi o'rganilgan tillardan biri ekani, shu bois undagi korpus va leksik resurslar boshqa turkiy tillar uchun transfer learning va qiyosiy tadqiqotlarda foydali bo'lishi mumkinligi alohida ta'kidlanadi [3]. Shu nuqtayi nazardan, o'zbek tili sintaktik analizatori uchun turk tilidagi yirik treebanklar va parser tajribalari nazariy hamda eksperimental manba vazifasini bajarishi mumkin. Biroq turkiy tillar bo'yicha tayyor resurslarni to'g'ridan-to'g'ri ko'chirib qo'llash ham yetarli emas. 2025-yilgi Turkic UD sharhi annotatsiya nomuvofiqliklari hanuz muhim muammo bo'lib qolayotganini ko'rsatadi. UD Turkic Groupning faoliyati ham aynan mavjud treebanklarni uyg'unlashtirishga qaratilgan. Demak, o'zbek tili bo'yicha dissertatsion tadqiqotimizda UT, UzUDT va TueCL resurslarini yagona

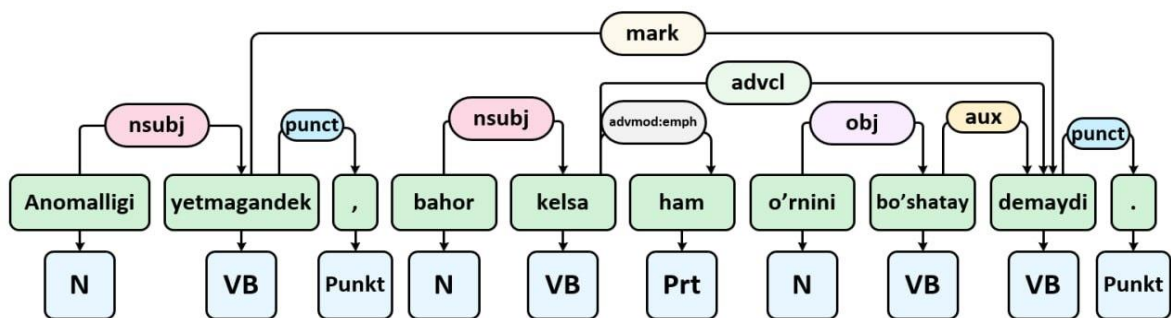
eksperimental tizimga birlashtirishdan oldin relatsiyalar tizimi, noaniq konstruksiyalar, tilga xos birliklar hamda yozuvni normallashtirish bo'yicha maxsus moslashtirish qoidalarini ishlab chiqish zarur. Aks holda, resurslar yig'indisi son jihatdan oshgan taqdirda ham, annotatsion nomuvofiqlik sabab model sifati sun'iy ravishda pasayishi mumkin. Bu xulosa rasmiy UD sahifalari va turkiy treebanklar bo'yicha moslashtirish tashabbuslariga tayangan ilmiy xulosadir.

Sintaktik tahlilning kompyuter lingvistikasidagi eng muhim metodologik masalalaridan biri bo'lib, tilning sintaktik tuzilmasi qaysi formalizm asosida ifodalanishi kerakligi masalasi hisoblanadi. Zamonaviy NLPda bu savolga javob, asosan, ikki asosiy usul orqali beriladi: *constituency parsing (birikma/bo'lakli yoki phrase-structure parsing)* va *dependency parsing (hokim-tobe munosabatlarga asoslangan parsing)*. Hozirgi ilmiy adabiyotlarda bu ikki yondashuv bir-biriga zid emas, balki sintaktik tuzilmani turli nuqtayi nazardan aks ettiruvchi o'zaro to'ldiruvchi formalizmlar sifatida baholanadi. Constituency tahlil daraxtlari gapning iyerarxik birikma tuzilishini aniq ko'rsatsa, dependency tahlil daraxtlari so'zlar orasidagi sintaktik munosabatlarni bevosita ochib beradi. So'nggi tadqiqotlarda ham aynan shu xususiyat sabab constituency va dependency tahlil daraxtlarini birgalikda modellashtirish g'oyasi ilgari surilmoqda [4].



1-rasm. Constituency parsing natijasi

So‘nggi yillarda ko‘plab kam resursli tillar uchun dependency treebanklar yaratilmoqda. Buning asosiy sabablaridan biri shundaki, dependency strukturalari intuitiv, amaliy jihatdan qulay va erkin so‘z tartibiga ega tillardagi lingvistik hodisalarni ifodalash uchun nisbatan moslashuvchan hisoblanadi. Dependency munosabatlari so‘zlar yoki undan kichik birliklar o‘rtasida aniqlanadigan sintaktik aloqalarni ifodalaydi. Bunda gap tarkibi hokim birliklar va argumentlarga ajratiladi, masalan, fe‘l va obyekt munosabatlari kabi. Dependency parsingning asosiy maqsadi leksik birliklar o‘rtasidagi ushbu bog‘lanishlarni aniqlash orqali ma‘lumotni tiklashdan iborat bo‘lib, bu ko‘pincha semantik interpretatsiya yoki sintaktik struktura shaklida namoyon bo‘ladi.



2-rasm. Dependency parsing natijasi

Parsing odatda tabiiy tilni qayta ishlashning dastlabki bosqichi sifatida qaraladi. Samarali va to‘g‘ri ishlaydigan parserlarni tayyorlash uchun zarur lingvistik ma‘lumotlar bilan annotatsiyalangan korpus talab etiladi. Annotatsiyaning qanchalik informativ yoki funksional bo‘lishi kerakligi haqida turli qarashlar mavjud. Annotatsiya tizimini loyihalash jarayonida bir qator cheklovlar va talablar hisobga olinadi. Jumladan:

- reprezentatsiya formatining qanchalik intuitiv va tabiiy ekani;
- undan ma‘lumot ajratib olishning qanchalik qulayligi;
- tabiiy tillarda uchraydigan hodisalarni qanchalik aniq va noaniqsiz ifodalay olishi.



Sintaktik parserlarni yaratish va o'qitishda annotatsiyalangan korpuslar asosiy lingvistik resurslardan biri hisoblanadi. Chunki parserlarning asosiy vazifasi gap tarkibidagi sintaktik bog'lanishlarni, grammatik munosabatlarni hamda hokim va tobe birliklar o'rtasidagi aloqalarni avtomatik ravishda aniqlashdan iboratdir. Bunday murakkab lingvistik vazifani bajarish uchun parser tizimlari oldindan sintaktik jihatdan belgilangan va tahlil qilingan ma'lumotlar asosida o'qitilishi zarur bo'ladi. Shu sababli annotatsiyalangan korpuslar parserlar uchun o'quv materialini bajaradi. Zamonaviy NLP tizimlarida, ayniqsa neyron tarmoqlar va transformer modellar asosidagi parserlarda annotatsiyalangan korpuslarning roli yanada ortib bormoqda. BERT, mBERT, RoBERTa kabi modellar yuqori aniqlikka ega bo'lishi uchun katta hajmdagi sintaktik annotatsiyalangan ma'lumotlar talab qilinadi. Shu sababli dependency treebanklar va boshqa turdagi annotatsiyalangan korpuslarni yaratish bugungi kunda kompyuter lingvistikasi va tabiiy tilni qayta ishlashning eng muhim ilmiy yo'nalishlaridan biri hisoblanadi. Sintaktik resurslar nuqtayi nazaridan ham o'zbek tili endi mutlaqo "resurssiz" emas, ammo hali "yetarli resursli" til darajasiga chiqmagan. UD Uzbek-UT [5] – o'zbek tili uchun birinchi UD treebank bo'lib, 500 gapdan tashkil topgan, yangilik va badiiy matnlarni qamrab oladi, uning POS va dependency relatsiyalari yarim avtomatik annotatsiyadan so'ng to'liq qo'lda tuzatilgan. O'zbek-UzUDT [6] esa to'liq manual gold-standard treebank sifatida 686 ta gap va taxminan 7.8 mingta tokenni o'z ichiga oladi, olti annotator ishtirokida yuqori interannotator agreement bilan yaratilgan. Uzbek-TueCL [4] esa 148 ta gapli, turkiy tillararo parallel grammatik misollar korpusining o'zbek tili qismi bo'lib, maxsus konstruksiyalarni qiyosiy o'rganish uchun muhimdir. Shu bilan birga, bu resurslarning o'lchami va janr qamrovi hali katta ko'lamli parser ekotizimi uchun yetarli emas. Demak, o'zbek tili uchun sintaktik model, annotatsiya qoidalari, algoritmi va dasturiy ta'minotni bir tizimda ishlab chiqish dolzarb ilmiy vazifa bo'lib qolmoqda.



Xulosa qilib aytganda, sintaktik tahlil NLP tizimlarining eng muhim komponentlaridan biri hisoblanadi. U til birliklari o'rtasidagi grammatik munosabatlarni aniqlash orqali matnni chuqur tushunish imkonini beradi. Mashinaviy tarjima, sentiment tahlili, savol-javob tizimlari, chatbotlar va generativ AI texnologiyalarining samaradorligi parserlarning sifatiga bog'liqdir. Universal Dependencies, treebanklar va transformer modellarining rivojlanishi sintaktik analizni yangi bosqichga olib chiqmoqda. O'zbek tilida ham katta hajmdagi annotatsiyalangan korpuslar va zamonaviy parserlarni yaratish NLP rivoji uchun muhim ilmiy yo'nalish hisoblanadi.

Foydalanilgan adabiyotlar ro'yxati

1. Marneffe M. C. De, Manning C. D., Nivre J., Zeman D. Universal dependencies. *Computational linguistics*, 47(2), 2021. Pp. 255-308.
2. Riezler S., Goldberg Y. Proceedings of the 20th SIGNLL conference on computational natural language learning. In Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, 2016, August.
3. Çağrı Toraman, Ahmet Kaan Sever and ets. TurkBench: A Benchmark for Evaluating Turkish Large Language Models. Proceedings of the Second Workshop Natural Language Processing for Turkic Languages (SIGTURK 2026), pages 126–154 March 29, 2026.
4. Gu Y., Hou Y., Wang Z., Duan X., Li Z. High-order joint constituency and dependency parsing. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 2024, May. Pp. 8144-8154.
5. https://universaldependencies.org/treebanks/uz_ut/index.html
6. https://universaldependencies.org/treebanks/uz_uzudt/index.html
7. https://universaldependencies.org/treebanks/uz_tuecl/index.html