

UDK: 811.512.133`42

KORPUS LINGVISTIKASIDA PARSERLARNING UMUMIY TAVSIFI

Abdullayeva Oqila Xolmo'minovna,
doktorant DSc
abdullayevaoqila@gmail.com
ToshDO'TAU

O'tkirova Fotima Baxtiyor qizi,
Kompyuter lingvistikasi yo'nalishi IV bosqich talabasi
otkirovafotima@gmail.com
ToshDO'TAU

Annotatsiya. Ushbu maqolada korpus lingvistikasining taraqqiyoti va jahonda yaratilgan parserlarning tavsifi, turlari, ularning ishlash prinsiplari, yaratilish maqsadi, hosilalar strukturasi hamda tilshunoslikdagi ahamiyati haqida so'z boradi. Rivojlanib borayotgan soha lingvistikasida parserlarning o'rni haqida batafsil ma'lumot beriladi.

Abstract. This article discusses the development of corpus linguistics and the description, types, operating principles, purpose of creation, structure of derivatives, and significance in linguistics of parsers created in the world. Detailed information is provided about the role of parser in the linguistics of the emerging field.

Аннотация. В данной статье рассматривается развитие корпусной лингвистики и описание, типы, принципы работы, цель создания, структура производных и значение созданных в мире парсеров в лингвистике. Предоставляется подробная информация о роли анализаторов в развивающейся отраслевой лингвистике.

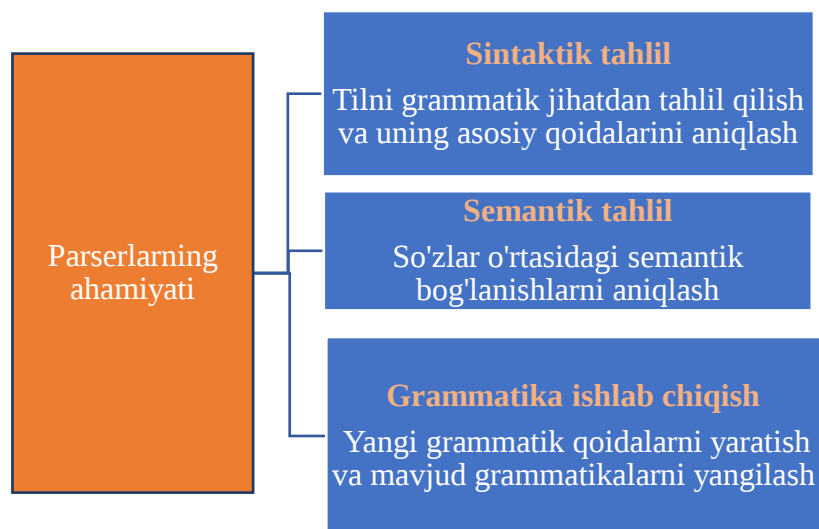
Kalit so'zlar: *Korpus lingvistikasi, korpus, parser, parser turlari, matn, tahlil, sintaktik tahlil, semantik tahlil, grammatik tahlil, treebank.*

XXI asrning global muammolaridan biri tabiiy tillarning milliy xususiyatini saqlab qolishdan iborat. Dunyo tillarining elektron korpuslarini yaratish va rivojlantirishda NLP hamda til texnologiyalariga doir tadqiqotlarni izchil ravishda olib borish dolzarb vazifaga aylandi[1]. Jahon tilshunosligida korpus lingvistikasining metodologiyasi yoki tilshunoslikning matn korpusini yaratish va tahlil qilish bilan shug'ullanuvchi bo'limi til tadqiqotlarini o'tkazishda keng qo'llaniladi. Korpus lingvistikasining rivojlanishi bilan matn materiallarining katta ma'lumotlari bazalarini izchil yaratish va kompyuter xotirasida saqlash mumkin

bo'ldi. Dunyoda tilshunoslikning taraqqiyoti talabi bilan jahon korpus tilshunosligida olib borilayotgan nazariy hamda amaliy izlanishlar bizning yurtimizda ham soha rivoji hamda mamlakat ravnaqi uchun kerakli va zarur nuqta ekani kundan kun ayon bo'lmoqda.

Korpus lingvistikasi – bu tilni matn korpusi orqali o'rganishning empirik usuli. Korpuslar – bu ma'lum bir til xilma-xilligini ifodalashga qaratilgan nutq yoki yozuv matnining muvozanatli, ko'pincha, qatlamli to'plamlari. Bugungi kunda korpuslar, odatda, mashina tomonidan o'qiladigan ma'lumotlar to'plami deya tadqiqotlarda eslatiladi. Ushbu soha XX asrning 60-yillarida kompyuter lingvistikasining bo'limi sifatida shakllangan tarmog'i bo'lib, kompyuter texnologiyalaridan foydalangan holda lingvistik korpuslarni (matnlar korpuslarini) qurish va ulardan foydalanishning umumiy tamoyillarini ishlab chiqish bilan shug'ullanadi. Lingvistik korpus yoxud til korpusi muayyan lingvistik muammolarni hal qilishga mo'ljallangan, maxsus saralangan va teglangan elektron shakldagi matnlar majmuasi hisoblanadi va pragmatik ahamiyatga ega eng yirik elektron baza sanaladi. Tilshunoslikda tilni o'rganish uchun foydalaniladigan matematik va kompyuter asosidagi metodlar kompyuter lingvistikasini tashkil etadi. Bu soha tilshunoslikning boshqa yo'nalishlaridan farqli ravishda empirik va statistik yondashuvlar bilan tahlil qilishni talab qiladi.

Soha mashinaga asoslangan tahlillarni amalga oshirishda bir qancha dasturlardan foydalanadi. Ana shunday soha texnologiyalarining asosiy vositalaridan biri parserlar (tilni qayta tahlil qilish dasturlari) hisoblanadi. Bunday dasturlar bir necha xil bo'lib, tilning tuzilishini, grammatikasini va semantik xususiyatlarini tushunish uchun kerak bo'lgan tizimlarni yaratadi. Korpus lingvistikasi matnlarni qayta tahlil qilishda empirik usullarga tayangani uchun korpusdagi mavjud so'z, gap va ular orasidagi o'zaro sintaktik aloqalarni aniqlashda parserlar asosiy vosita sanaladi. Korpuslar – tilni qayta real ishlatishdagi barcha xususiyatlarini o'zida jamlagan matnlar majmuyi. Tahlil qiluvchi dasturlar matnlarni sintaktik analiz qilib, grammatik modellarni qayta quradi. Korpus lingvistikasi yordamida tilshunoslar grammatik tuzilmalar, so'z birikmalari, kollokatsiyalar va boshqa til xususiyatlarini tahlil qilishadi. Matnlarning grammatik tuzilishini aniqlashda parserlar muhim dastur sanaladi. Bunday dasturlar, ko'pincha, quyidagi maqsadlarda qo'llaniladi:

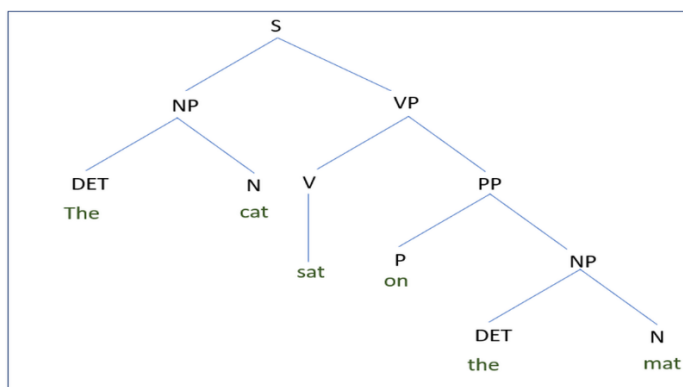


1-rasm. Parserlarning ahamiyati

Parserlarning turi va ularning tavsifi

Matnni grammatik jihatdan tahlil qilib, uning sintaktik strukturasi aniqlovchi dasturiy ta'minot vositalariga **parserlar** deyiladi. Bugungi kunda parserlarning quyidagi turlari mavjud:

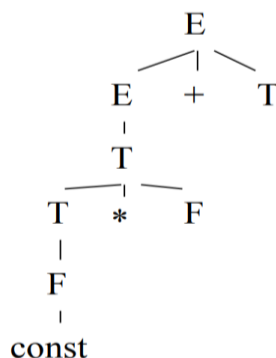
1. Sintaktik parserlar kiritilgan gaplarning sintaksis strukturasi aniqlaydi va ushbu strukturaga mos keluvchi sintaktik sxema quradi. Dastlab, jumalarni sintaktik jihatdan tahlil qiladi. Matnlarni ma'lum bir til grammatikasiga moslashtirib, iboralar va tilda gaplarni tuzish uchun qo'llaniladigan yaxlit birliklarni ajratib oladi. Sintaktik tahlilning maqsadi qanday qilib tarkibiy qismlar gaplarni shakllantirish uchun ma'lum bir sintaktik tartibda joylashtiriladi, degan savolga javob berish hisoblanadi. Bunday tahlil dasturlarida gaplardagi mavjud so'zlarning o'zaro munosabati hamda bog'liqligi yakuniy natijasini amalga oshirishda muhim rol o'ynaydi. Zamonaviy tilshunoslikda parsing, asosan, kompyuter yordamida sintaktik tahlil qilishni anglatadi(2-rasm).



2-rasm. Sintaktik parser natijasi

Bu turdagi parserlar ikki turga bo'linadi:

Top-down (yuqoridan pastga) parserlar – gaplarni grammatik qoidalarga tayanib tahlil qiladi. Bunday parserlar yuqoridan pastga usulidan foydalanadi hamda tahlil sxemasi boshlang'ich belgidan ildiz sifatida pastki qismdagi terminal belgilarga qarab o'sib borishi tarzida shakllanadi(3-rasm). Oddiy tillar bilan ishlashda bunday tahlilni amalga oshirish va tushunish pastdan yuqoriga qarab shakllangan tahlildan osonroq.

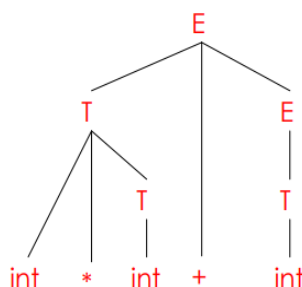


3-rasm. Top-down parserning tahlil sxemasi

Bunday tahlil qilishning yana-da samarali varianti ham bor bo'lib, u **LL-Parsing** deb ataladi. Bunday tahlilda *Lookhead* belgisi hisobga olinib, keyingi qoidani qo'llash kerakligini oldindan aytib turish imkoni bo'ladi. LL sinfidagi parserlar top-down parserlar singari tahlil jarayonida ortga qaytishi shart bo'lmaydi.

Rekursiv Descent Parserlar Top-down parserlarning yana bir turi bo'lib, o'ziga xos yuqoridan pastga yo'natirilgan tahlil dasturi sanaladi. Bu turdagi tahlil usuli rekursiv ravishda ishlaydi va qoidalarni bosqichma-bosqich amalga oshiradi.

Bottom-up (pastdan yuqoriga) parserlar – berilgan matndagi so'zlardan tortib, grammatik qoidalar yordamida yaxlit sintaktik struktura hosil qiladi. Dastlab, dastur matndagi jummalarni so'z birikmalariga ajratib oladi. Keyin ushbu birikmalar umumiy grammatik tuzilishga asoslangan holda guruhlanadi. Oxirida butun jumlaning sintaktik strukturasi hosil qilinadi(4-rasm).

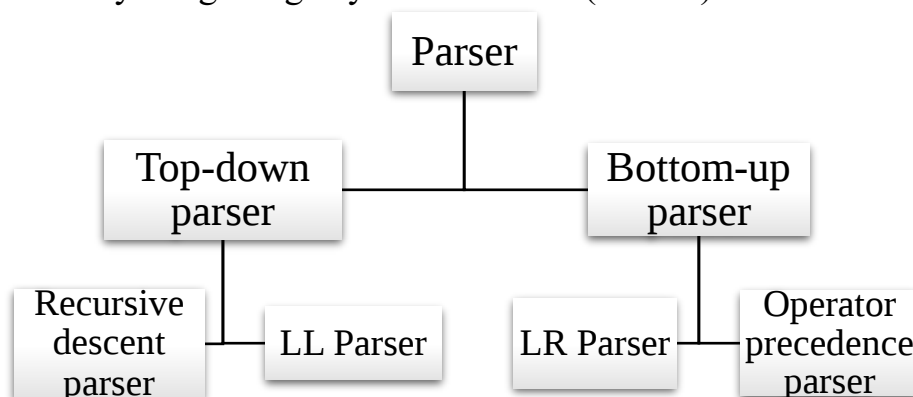


4-rasm. Bottom-up parser sxemasi

Bottom-up parserlarning **LR Parser (Left to right)** va **Operator Precedence Parser** kabi 2 turi mavjud. LR Parserlar jumlaning chapdan o'ngga qarab tahlil qiladi. Bunday dastur tahlili bilan murakkab bo'lgan grammatik qoidalarni tushunish

mumkin bo'лади. Tez ishlaydi hamda kuchli parser hisoblanadi. Operator Precedence Parser, asosan, operatorlarning ustuvorlik darajasini aniqlash uchun qo'llaniladi. Bunday dastur turlari oddiyroq bo'lgan grammatik qoidalarni tahlil qilish uchun juda samarali deb baholanadi.

Bugungi kunda jahon tilshunosligida yaratilgan sintaktik tahlil dasturlari tilshunoslikka ko'plab tahlil usullari va grammatik strukturalarni taqdim etdi. Bu kabi aqlli parserlarning yaratilishi o'ziga xos bo'lgan tahlil metodlari tufayli ko'plab istisnoli sintaktik vaziyatlarga to'g'ri yechimlar berdi(5-rasm).



5-rasm. Parserlar klassifikatsiyasi

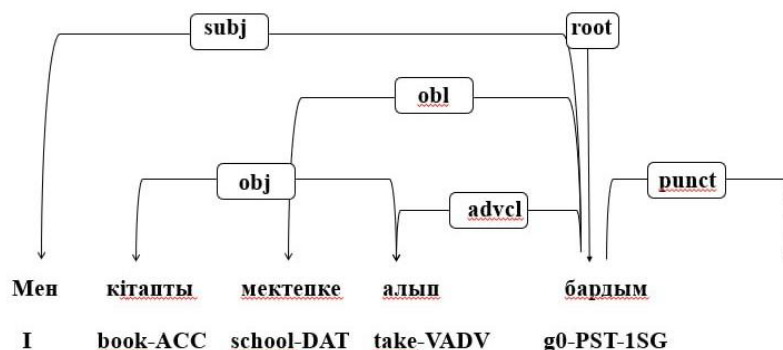
2. **Semantik Parserlar** matnning ma'no jihatidan tahlilini amalga oshiradi. Ular gaplardagi mavjud leksemalarning o'zaro bog'liqligini aniqlash maqsadida ishlab chiqiladi. Semantik parserlar so'zlarning ma'nosini matnning umumiy mazmuniga aloqador holatda aniqlaydi. Masalan, “Dugonam gulni hidladi” va “Dugonam gulni uzdi” jumlaridagi *gul* so'zi har ikki gapda ham turli rol o'ynagan. Semantik parser yordamida bu farqni aniqlash osonlashadi. Bir qancha muammolarga qaramasdan, bugungi kunda yuqori aniqlikda ishlovchi semantik parserlar yaratilgan. Bu kabi muammo turlaridan ayrimlari tabiiy tildagi gaplarning noaniq bo'lishi, gaplar murakkab tuzilmalar va munosabatlarga ega bo'lishidir.

3. **Stoxastik va Statik parserlar.** Bu kabi parser turlarini ishga tushirish uchun bir nechta inson omili bilan bog'liq jarayonlar talab etiladi. Jumladan, katta hajmdagi ma'lumotlar to'plamidan foydalanib, dasturga muayyan matnni to'g'ri tahlil qilish o'rgatiladi hamda eng ehtimoliy modellar asosida tahlil qila boshlaydi. Bu usulda yaratilgan tahlilchi dasturlar bir necha million misollarni tahlil qilib, ma'lum grammatik strukturalarni o'rganib chiqadi.

Statik parserlar avvaldan belgilangan grammatik qoidalarga asoslangan bo'lib, ayni matnni aniq sintaktik tuzilishga ajratadi. Bunday tahlil dasturlarida barcha grammatik qoidalar oldindan belgilanadi.

4. **Dependency (bog'liqlik) parserlarda** har bir so'z boshqa biror so'zga bog'lanadi va bu bog'lanish tizimi biror “root” so'zi atrofida tashkil topadi. Root so'zi – gapdagi asosiy so'z bo'lib, barcha boshqa so'zlar shu asosda bog'lanadi. Masalan, “Men kitob o'qidim” degan jumlada “o'qidim” – root so'z bo'lib, “men”

va “kitob” so‘zlari unga bog‘langan. Natijada bog‘lanishlar yo‘naltirilgan chiziqlar shaklida ifodalanadi. Har bir bog‘lanish bir so‘zni boshqasiga bog‘laydi va bu bog‘lanish yo‘naltirilgan bo‘ladi. Ya’ni, bir so‘z ikkinchisiga bog‘lanib, uni boshqaradi. Bunday munosabatlar grammatik jihatdan gapning to‘g‘riligini tekshiradi. Bunday tahlil turida gapning strukturasi tasvirlash uchun turli diagrammalar (dependency trees) ishlatiladi. Har bir so‘z node (tugun) sifatida ifodalanadi va ular o‘rtasidagi bog‘lanishlar yo‘naltirilgan chiziqlar bilan ko‘rsatiladi. Diagrammalarni ko‘rish orqali tilshunoslar va kompyuter tizimlari grammatik tuzilmalarni osongina tahlil qilishi mumkin(6-rasm).



6-rasm. Dependency parsing natijasi

Xulosa qilib aytganda, Dependency parsing (bog‘lanish tahlili) va, umuman, parser dasturlari tabiiy tilni qayta ishlash (NLP) tizimlarining, masalan, savol-javob tizimlari yoki mashina tarjimai kabi ko‘plab sohalarida muhim bosqich hisoblanadi. o‘zbek tili morfologik jihatdan boy va grammatik jihatdan murakkab til bo‘lib, uni avtomatik tarzda qayta ishlashni yanada qiyinlashtiradi. O‘zbek tiliga oid NLP vositalari va resurslarining cheklanganligi bu vazifani yanada murakkablashtiradi. Qoidalarga asoslangan hamda ma’lumotlarga asoslangan chuqur o‘rganish (deep learning) modellari parserlarni ishlab chiqish va amaliyotda qo‘llashda natijalarni ko‘rsatmoqda. Biroq, chuqur o‘rganishga asoslangan tizimlar yaxshi natijalarga erishish uchun katta hajmdagi ma’lumotlarga muhtoj bo‘ladi. Bu boradagi tadqiqotlar chuqur o‘rganilishiga ehtiyoj mavjudligini ko‘rsatadi.

Foydalanilgan adabiyotlar:

1. Abdurahmonova N., Urdishev K. Corpus based teaching Uzbek as a foreign language. / Journal of Foreign Language Teaching and Applied Linguistics (J-FLTAL). 2019.
2. Salvatella M.L. Parsing and Evaluation. Improving dependency grammars accuracy. PhD. diss – Universitat de Barcelona, 2016.
3. Ruket Cakici. Wide Coverage parsing for Turkish. –M., 2009.



4. Collins Michael, James Brooks. Prepositional attachment through a backed-off model. –Cambridge., 1995.
5. Foth K., Menzel W. Learning the constraints weights of a dependency grammar using genetic algorithms // 3th International conference on Domain Decomposition Methods, 2001. –P 243-247.
6. Abjalova M. Korpus lingvistikasi/uslubiy qo'llanma. –Toshkent, 2022.
7. <https://www.geeksforgeeks.org/types-of-parsers-in-compiler-design/>