

O'ZBEKCHA SO'ZLARNING GRAMMATIK TAHLIL QILISH ALGORITMLARI

Qarshiyev Abduvali

f.-m.f.n., professor. Muhammad al Xorazmiy nomidagi
Toshkent axborot texnologiyalari universiteti Samarqand filiali.

Karimov S.A.

Filologiya fanlari doktori, professor
Samarqand davlat universiteti.

Tursunov M.S.

Muhammad al Xorazmiy nomidagi
Toshkent axborot texnologiyalari universiteti
Samarqand filiali, ilmiy izlanuvchi.

Annotatsiya. Ushbu maqolada o'zbekcha so'zni grammatik tahlil qilish algoritmlari, usullarni tavsiflash va kodlashga umumiy yondashuv muhokama qilinadi. Ishda so'zni grammatik tahlil qilishda dastavval so'z shaklidan so'zning asosini va qo'shimchalar qismini ajratib olish. Undan so'ng qo'shimchlarni bo'laklarga ajratish hamda so'z asosida mos turkumni aniqlash hamda qo'shimchalar tahlilini bajarishdan iborat. Ishlab chiqilgan algoritm asosida nafaqat so'zlarni balki matnlarni grammatik tahlil qilish uchun foydalansa bo'ladi. Bu esa korpus tilshunosligida ham muhim ahamiyat kasb etadi.

Kalit so'zlar: so'z shakli, so'z asosi, model, algoritm, ma'lumotlar bazasi, lug'at, grammatika, morfologik tahlil.

Аннотация. В данной статье рассматриваются алгоритмы грамматического анализа узбекских слов, описание методов и общий подход к кодированию. В работе при грамматическом анализе слова прежде всего необходимо отделить от словоформы основу слова и суффиксальную часть. После этого он состоит из разделения наречий на части и определения на основе слова соответствующей категории и выполнения анализа наречий. На основе разработанного алгоритма его можно использовать для грамматического анализа не только слов, но и текстов. Это также важно для корпусной лингвистики.

Ключевые слова: словоформа, словесная база, модель, алгоритм, база данных, словарь, грамматика, морфологический анализ.

Annotation. This article discusses algorithms for grammatical analysis of Uzbek words, description of methods and a general approach to coding. At work, in the grammatical analysis of the word, first of all, it is necessary to separate the base of the word and the suffix part from the word form. After that, it consists of dividing the adverbs into parts and determining the appropriate category based on the word and performing the analysis of the adverbs. Based on the developed algorithm, it can be used for grammatical analysis of not only words, but also texts. This is also important in corpus linguistics.

Keywords: word form, word base, model, algorithm, database, dictionary, grammar, morphological analysis.

O'tgan asrning 90-yillarida morfologik tahlil *generativ model*, *paradigmatik model*, *ikki tarkibli morfologik model* [Mahlov C., Piotrowski M., 2009, 15] kabi metodlarga asoslanilgan. O'ttiz yildan oshiqroq vaqt mobaynida kompyuter morfologiyasi uchun turli texnologiyalarni yaratishga doir tadqiqotlar olib borilmoqda. Tillar uchun morfologik tahlil qilishning o'nlab algoritmlari bo'lishi tabiiy hol. Algoritmning bir yo'nalishi so'z oxiridagi harfiy birikmalarni qamrab oluvchi ma'lumotlar bazasidan foydalanadi. Bunda ehtimollik darajasiga ko'ra, lingvistik resursdan olingan natijalar statistik tahlil qilinadi. Avtomatik usulda morfologik tahlil qilish jarayoni quyidagi turlarga ko'ra tasniflanadi [A.Po'latov, S.Muhamedova, 2008]:

1. Mantiqiy ko'paytirish metodi orqali morfologik analiz.
2. Lug'atsiz jadval orqali morfologik analiz.
3. So'zlar asosiga asoslangan morfologik tahlil (so'zlar asosiga tayanilgan morfologik tahlil Yevropa tillari uchun moslashtirilgan. Unda qo'shimcha yordamchi jadvallardan foydalaniladi).

4. So'zshakllarga (morfotaktik qoidalarga ko'ra) ko'ra morfologik analiz.

Yuqoridagilardan so'zshakllariga ko'ra morfologik tahlil qilish usuli keng qo'llanadi. Ushbu algoritmgga muvofiq so'zlar morfemalarga ajratilganda qidirilayotgan so'z lug'atdan topilmaslik ehtimoli kuzatiladi. Bu holatda o'ngdan chapga qarab tahlil qilish jarayonida ko'makchi morfemalarning turi va modeliga qarab uning qaysi so'z turkumiga oidligi aniqlanadi. U holda o'sha so'zning qaysi so'z turkumiga tegishli ekanligi belgilanadi [Tursunov M.S., Qarshiyev A.B., 2019, 290-291]. So'z shakllarni tahlil qilish jarayonida quyidagi bosqichlar amalga oshiriladi:

1. So'zshakl morfem segmentlarga ko'ra zanjirsimon birikadi.
2. So'zshakl morfem segmentlar zanjiri sifatida tartiblanmagan yangi elementlar to'plami - morfemalarga almashtiriladi. So'z shakllar morfemalar to'plami sifatida ma'lumotlar yoziladi.
3. Ushbu ma'lumot so'z shakllari haqidagi yakuniy natija bilan qayta to'ldiriladi.
4. Mantiqiy ko'paytirish metodi orqali morfologik tahlil qilish flektiv tillar uchun foydalaniladi

O'zbek tilidagi so'z shakllarini grammatik asosini aniqlash muhim ahamiyatga ega.

O'zbek tili so'z shakllarini grammatik tahlil qilish morfotaktik qoidalarga asoslangan holda amalga oshiriladi.

O'zbek tilidagi so'z shakllarini morfologik asosini aniqlashda so'zning strukturasi muhim ahamiyat kasb etadi. Shu munosabat bilan quyidagi matematik ifodalar orqali tasvirlangan lingvistik qoidalar asosidagi affiksasiya usuli bilan yasalgan so'z shakllarining morfemik modellari taklif etildi [Baqayev I.I., 2021, 11]:

$$w_{sy} = K + \left\{ \sum_{l=1}^m sx_l \right\} + \left\{ \sum_{i=1}^n af_i \right\}.$$

Bu yerda K- sodda tub so'z (o'zak), sx - so'z yasovchi affiks (suffixs), af - affiksar, $\Sigma - 0,1, \dots n$ - ketma-ket konkatenasiya operatsiyasi, l - suffiksar miqdori indeksi, i - affiksar miqdori indeksi w_{sy} - sodda yasama so'z. Fikrimizni dalillash uchun ot so'z turkumidagi «gul-don-lar-imiz-ni» so'z shaklini tahlil qilamiz. Bu misolda K- «gul», sx - «-don», af - «-lar», «-imiz», -ni «, «gul-don-lar-imiz-ni» - w_{sy} so'z shakli [4].

Bunda so'z shaklidan so'zning ajratib olinadi, hamda keyingi qadamda qo'shimchalardan tashkil topga bo'lagi tahlil qilinadi. So'z shaklini grammatik tahlil qilish jarayonida albatta

ma'lumotlar bazasiga murojaat qilinadi. Biz ilmiy ishimiz uchun ma'lumotlar bazasida ikkita jadvaldan foydalanamiz: 1) so'zlar asosi lug'ati, 2) qo'shimchalar lug'ati (1-rasm). So'zlar asosi lug'ati 3ta maydondan iborat: 1) tartib raqam (So'z asosining IDsi), 2) So'z asosi, 3) So'z asosining turkumi. Qo'shimchalar lug'ati ham 3ta maydondan iborat: 1) tartib raqam (qo'shimchani IDsi), 2) Qo'shimchani nomi, 3) Qo'shimchani grammatik ma'noga ega izohi.

ID	Asos	Turkum
1	kitob	ot
2	qizil	sifat
...	...	...

a)

ID	Qo'shimcha	Izoh
1	lar	ko'plik
2	im	o'zlik
...	...	...

b)

1-rasm. a) So'zlar asosi lug'ati, b) Qo'shimchalar lug'ati.

So'zlarni ma'lumotlar bazasidan o'qib olish uchun ularning simvollar ketma- ketligiga amal qilinadi. Belgilash o'sha so'z yaxlit birlik sifatida aniqlanmagunicha bu holat takrorlanaveradi. Morfologik tahlil uchun komponent tahlil metodidan ham foydalanish mumkin. Unga ko'ra so'zning uzunligi aniqlanadi. Shu tarkibdagi simvollar soniga qarab ma'lumotlar bazasidan qidiriladi. So'zning asosini topish va uni grammatik tahlilini qilishda ishlab chiqilgan algoritm quyidagi ketma ketlikdan iborat:

1-qadam. Dastur ishini boshlash.

2-qadam. <input> tegiga so'z shakli kiritiladi.

3-qadam. Input tegiga kiritilgan so'z shakli "suz" nomli o'zgaruvchiga ta'minlanadi.

4-qadam. "suz" o'zgaruvchining qiymati, "suzl" nomli o'zgaruvchiga ta'minlanadi, ya'ni so'zning nusxasi yaratiladi.

5-qadam. "l" o'zgaruvchisiga "suzl" o'zgaruvchisi qiymatining uzunligi, ya'ni so'z shaklining belgilar soni ta'minlanadi.

6-qadam. "suzl" o'zgaruvchisiga ta'minlangan so'z shaklidan so'z asosini aniqlash algoritmi boshlanadi. Bunda berilgan so'z shakli ma'lumotlar bazasidan qidiriladi. Agar mavjud bo'lsa 11-qadamga o'tiladi, aks holda so'zning oxiridan boshlab bittadan belgi kesib tashlash tartibida lug'atdan qayta qidirish shaklida siklimiz davom etadi. Bunda sikl aylanishi toki so'z asosi topilgunga qadar yoki "suzl" o'zgaruvchining qiymati bo'sh bo'lib qolgunga qadar davom etadi. "suzl" o'zgaruvchisining qiymati bo'sh bo'lib qolsa, u holda so'z shaklidan uning asosini aniqlashni imkoni bo'lmagan, va "so'z asosini aniqlab bo'lmadi" degan xabarni berib, 12-qadamga o'tiladi. "suzl" o'zgaruvchisidagi so'z shaklidan uning asosi topilsa siklimiz rost qiymatni olib, davom etish to'xtatiladi va 7-qadamga o'tiladi.

7-qadam. "suzl" o'zgaruvchisidagi so'z shaklidan uning asosi o'zgaruvchining o'ziga ta'minlangan bo'ladi. "suz" o'zgaruvchisi qiymatidan "suzl" o'zgaruvchisi qiymati kesib olinib, "qushl" o'zgaruvchisiga ta'minlanadi, ya'ni "suzl" o'zgaruvchisida so'z shaklining asosi, "qushl" o'zgaruvchisida so'z shaklining qo'shimchalar qismi yaratiladi.

8-qadam. "qushl" o'zgaruvchisi qiymati sifati so'z shaklini qo'shimchalar qismini olgan va undan qo'shimchalarni aniqlash algoritmi boshlanadi. Bunda qo'shimchalar lug'atidan qo'shimchani izlaydi. Agar topilsa 11-qadamga o'tiladi, aks holda qo'shimchalarni Grammatik tahlil qilish uchun bo'laklash boshlanadi

9-qadam. "qushl" o'zgaruvchisi qiymatining birinchi belgisidan boshlab "qo'shimchalar lug'atidan qidiriladi. Agar lug'atdan topilmasa keyingi belgini avvalgi belgiga qo'shgan holatda

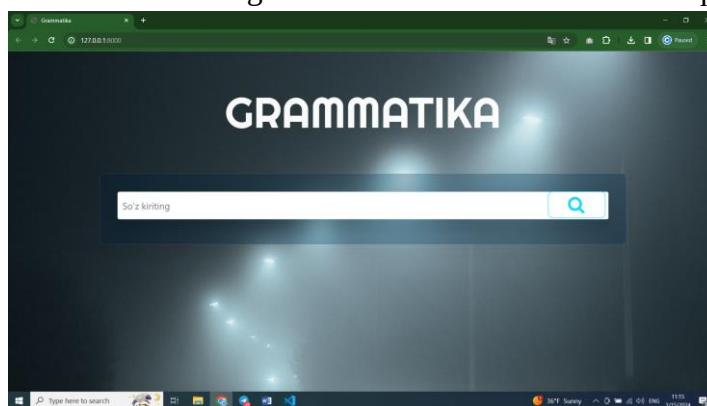
lugʻatdan qidirish davom etadi. Bu amal toki *“qushl”* ning qiymatidagi belgilarning oxirigacha davom etadi. Agar topilmasa 12-qadamga oʻtiladi, aks holda dastur ishini davom ettiradi.

10-qadam. *“qushl”* qoʻshimchalar ketma-ketligidan iborat satrda birinchi belgidan boshlab tekshirish (9-qadamda) davomida birinchi qoʻshimcha aniqlanganda ushbu qoʻshimcha va uning izohini massivga yozadi va *“qushl”* oʻzgaruvchisining qiymatidan aniqlanga qoʻshimcha kesib tashlanadi. Masalan: -larining, -lar+-i-ning. Bunda -lar qoʻshimchasi aniqlanganda *“qushl”*ning qiymati -ining boʻladi va 9-qadamdagi algoritimga qaytadi.

11-qadam. Bunda morfolig tahlil qilingan maʼlumotlar soʻz shaklining asosi, turkumi, qoʻshimchalar roʻyxati va ularning izohi foydalanuvchiga taqdim etiladi.

12-qadam. Dastur ishini tugatadi.

Ushbu algoritm asosida morfologik razmetkalash dasturi ishlab chiqilgan.



2-rasm. Dasturning asosiy interfeysi

Soʻz shaklini Grammatik razmetkalash uchun soʻz shakli kiritiladi. Misol sifatida *“kitoblarga”* soʻzni tahlil qilaylik. Bunda 3-rasmda koʻrsatilgandek soʻz shaklini Grammatik tahlil qilib beradi.

Grammatik tahlil xulosalarini quyidagicha koʻrish mumkin:

Soʻz shakli – *“kitoblarga”*;

Soʻz asosi – *“kitob”*;

Soʻz turkumi – *“ot soʻz turkumi”*;

Qoʻshimchalar roʻyxati – *“-lar, -ga”*;

Qoʻshimchalarga berilgan izoh – *“koʻplik, joʻnalish kelishigi”*;

Bundan oʻrganayotgan kalit soʻzimiz ot soʻz turkumiga taalluqligi, unda koʻplik va joʻnalish kelishiklaridagi qoʻshimchalar toʻgʻrisidagi lingvistik maʼlumotlar olingan.



3-rasm. Grammatik maʼlumotlar taqdimoti

Adabiyotlar ro'yxati:

1. Mahlov C., Piotrowski M., JSLIM – Computational Morphology in the Framework of the SLIM, Theory Language / State of Art in Computational Morphology, Zurich, 2009, 15-p.
2. A.Po'latov, S.Muhamedova. Kompyuter lingvistikasi. -T.: 2008.
3. M.Tursunov, A.Qarshiyev. O'zbekcha matnlarni morfologik va leksik tahlil qilish algoritmlari va dasturlari / International scientific conference "Actual problems of applied mathematics and information technologies. –Toshkent:2019, 290-29.
4. Baqayev I.I., O'zbek tili so'z shakllarini morfologik tahlil qilish modellari va algoritmlari / avtoreferat, Toshkent: TATU, 2021, b. 11