

O'ZBEK TILIDA FE'L ANALITIK SHAKLLARINI AVTOMATIK TANIB OLIISH MUAMMOLARI

Nurova Firuza Kamolovna,

Buxoro davlat universiteti O'zbek tilshunosligi va jurnalistika kafedrasida o'qituvchisi

Annotatsiya. Mazkur maqolada o'zbek tilida fe'l analitik shakllarini avtomatik tanib olish muammolari kompyuter lingvistikasi nuqtayi nazaridan tahlil qilinadi. Tadqiqotda fe'l analitik shakllarining lingvistik tabiati, ularning morfologik va sintaktik xususiyatlari hamda tabiiy tilni qayta ishlash tizimlarida uchraydigan muammolar yoritilgan. Xususan, analitik shakllarni avtomatik aniqlash jarayonida yuzaga keladigan omonimiya, yordamchi fe'llarning ko'p funksiyaliligi, analitik va sintaktik konstruksiyalar o'rtasidagi chegaralarning noaniqligi, korpus ma'lumotlarining yetarli emasligi kabi masalalar ko'rib chiqilgan. Shuningdek, analitik shakllarni aniqlashda qoidalarga asoslangan, statistik va neyron tarmoqlar asosidagi yondashuvlarning imkoniyatlari tahlil qilinadi. Tadqiqot natijalari o'zbek tili morfoanalizatorlari, sintaktik analizatorlar va mashina tarjimai tizimlarini takomillashtirishda muhim ahamiyat kasb etadi.

Kalit so'zlar: fe'l analitik shakllari, kompyuter lingvistikasi, morfologik analiz, morfoanalizator, tabiiy tilni qayta ishlash, yordamchi fe'llar, omonimiya, disambiguatsiya, korpus lingvistikasi, CRF, BERT, mashina tarjimai, sintaktik tahlil, o'zbek tili.

Аннотация. В данной статье анализируются проблемы автоматического распознавания аналитических форм глаголов в узбекском языке с точки зрения вычислительной лингвистики. Исследование охватывает лингвистическую природу аналитических форм глаголов, их морфологические и синтаксические свойства, а также проблемы, встречающиеся в системах обработки естественного языка. В частности, рассматриваются такие вопросы, как омонимия, возникающая в процессе автоматического распознавания аналитических форм, многофункциональность вспомогательных глаголов, неоднозначность границ между аналитическими и синтаксическими конструкциями, а также недостаток корпусных данных. Также анализируются возможности подходов к распознаванию аналитических форм, основанных на правилах, статистических методах и нейронных сетях. Результаты исследования имеют большое значение для совершенствования узбекских языковых морфоанализаторов, синтаксических анализаторов и систем машинного перевода.

Ключевые слова: аналитические формы глаголов, вычислительная лингвистика, морфологический анализ, морфоанализатор, обработка естественного языка, вспомогательные глаголы, омонимия, разрешение неоднозначности, корпусная лингвистика, CRF, BERT, машинный перевод, синтаксический анализ, узбекский язык.

Abstract. This article analyzes the problems of automatic recognition of analytical verb forms in the Uzbek language from the perspective of computational linguistics. The study covers the linguistic nature of analytical verb forms, their morphological and syntactic properties, and problems encountered in natural language processing systems. In particular, it addresses issues such as homonymy arising during the automatic recognition of analytical forms, the multifunctionality of auxiliary verbs, the ambiguity of the boundaries between analytical and syntactic constructions, and the lack of corpus data. The potential of rule-based, statistical, and neural network-based approaches to recognizing analytical forms is also analyzed. The results of the study are of great importance for improving Uzbek language morphoanalyzers, parsers, and machine translation systems.

Keywords: analytical verb forms, computational linguistics, morphological analysis, morphoanalyzer, natural language processing, auxiliary verbs, homonymy, disambiguation, corpus linguistics, CRF, BERT, machine translation, syntactic analysis, Uzbek language.

Kirish. Kompyuter lingvistikasining zamonaviy yo‘nalishlaridan biri tabiiy tilni qayta ishlash (Natural Language Processing – NLP) texnologiyalarini ishlab chiqish hisoblanadi. Mazkur yo‘nalishda matnlarni avtomatik tahlil qilish, mashina tarjimasini, axborot qidirish tizimlari, nutqni tanish va sintez qilish kabi masalalar alohida ahamiyat kasb etadi. Ushbu vazifalarni samarali bajarish uchun til birliklarining grammatik xususiyatlarini aniq modellashtirish talab etiladi. O‘zbek tili uchun bunday muammolardan biri fe‘l analitik shakllarini avtomatik tanib olish masalasidir.

Turkiy tillar, jumladan o‘zbek tili morfologik jihatdan agglutinativ tizimga ega bo‘lib, grammatik ma‘nolar ko‘pincha qo‘shimchalar va yordamchi birliklar vositasida ifodalanadi. Fe‘l tizimida esa grammatik ma‘nolarni ifodalashda analitik shakllar muhim o‘rin tutadi. Analitik shakllar bir necha komponentlardan tashkil topib, yaxlit grammatik ma‘noni ifodalaydi. Masalan, *yozib qo‘ydi, kelib qoldi, aytib yubordi, ishlab turibdi, kelgan edi* kabi birliklar tarkiban ikki yoki undan ortiq komponentdan iborat bo‘lsa-da, nutqda yagona predikativ ma‘no hosil qiladi.

Analitik shakllarni avtomatik aniqlash va ularni boshqa sintaktik konstruksiyalardan farqlash o‘zbek tili morfoanalizatorlari va sintaktik analizatorlarini yaratishda dolzarb muammolardan biri hisoblanadi.

Adabiyotlar tahlili. O‘zbek tilida fe‘l analitik shakllarini avtomatik tanib olish muammosi kompyuter lingvistikasi va tabiiy tilni qayta ishlash sohasida dolzarb masalalardan biridir. A.A. Abduazizov tilshunoslik nazariyasida grammatik birliklarning tizimli tahlili muhimligini ta’kidlaydi.[1] A. Hojiev esa o‘zbek tilida ko‘makchi fe‘llar va to‘liqsiz fe‘llarning analitik shakllar hosil qilishdagi rolini ilmiy asoslab bergan.[7. 8]

Shuningdek, N. Abdullayeva va boshqa tadqiqotchilar korpus lingvistikasi va NLP tizimlarida til birliklarini avtomatik qayta ishlash muammolarini yoritgan.[2.4] A. Norov ishlarida esa kompyuter lingvistikasining nazariy asoslari va morfoanaliz jarayonlari tahlil qilingan.[6]

Umuman olganda, mavjud ilmiy adabiyotlar fe‘l analitik shakllarini avtomatik tanib olishda omonimiya, yordamchi fe‘llarning ko‘p funksiyaliligi va korpus yetishmovchiligi asosiy muammolar ekanini ko‘rsatadi. Shu bilan birga, ushbu yo‘nalishda qoidalarga asoslangan, statistik va neyron tarmoqlar yondashuvlarini uyg‘unlashtirish zarurligi ta’kidlanadi.

Tadqiqot metodologiyasi. Mazkur tadqiqotda o‘zbek tilida fe‘l analitik shakllarini avtomatik tanib olish muammolarini yoritishda tavsifiy-analitik, qiyosiy va korpus lingvistikasi metodlaridan foydalanildi. Tavsifiy-analitik metod yordamida analitik fe‘l shakllarining lingvistik xususiyatlari, ularning morfologik va sintaktik tuzilishi tahlil qilindi.

Qiyosiy metod asosida analitik va sintaktik konstruksiyalar o‘rtasidagi farqlar hamda yordamchi fe‘llarning mustaqil va grammatik vazifalari solishtirildi. Korpus lingvistikasi

metodi esa mavjud matnlar asosida analitik shakllarning qo‘llanish holatlarini kuzatish va tahlil qilish imkonini berdi.

Shuningdek, tadqiqotda kompyuter lingvistikasi yondashuvlari — qoidalarga asoslangan, statistik va neyron tarmoqlar asosidagi modellar imkoniyatlari ham nazariy jihatdan ko‘rib chiqildi.

Tahlillar va natijalar. O‘zbek tilshunosligida analitik shakl deganda ikki yoki undan ortiq mustaqil yoki yordamchi komponentlarning birlashuvi natijasida yuzaga kelgan va yagona grammatik ma‘noni ifodalovchi birlik tushuniladi. Fe‘l analitik shakllari ko‘pincha ravishdosh, sifatdosh yoki harakat nomi shakllarining yordamchi fe‘llar bilan birikishi orqali hosil bo‘ladi.

Masalan: yozib yubormoq; olib kelmoq; aytib qo‘ymoq; ishlab turmoq; kelgan ekan; borayotgan edi.

Bu shakllarda asosiy leksik ma‘no yetakchi fe‘l tomonidan ifodalansa, grammatik va modal ma‘nolar yordamchi fe‘l orqali yuzaga chiqadi.

Analitik shakllar quyidagi grammatik ma‘nolarni ifodalashi mumkin:

1. Harakatning tezligi;
2. Davomiylilik ma‘nosi;
3. Boshlanish ma‘nosi;
4. Tugallanganlik ma‘nosi;
5. Kutilmaganlik ma‘nosi;
6. Modal ma‘nolar;

Shuning uchun ularni avtomatik aniqlash jarayonida nafaqat morfologik, balki semantik va sintaktik belgilar ham hisobga olinishi lozim.

Avtomatik tanib olishning nazariy asoslari. Kompyuter lingvistikasida analitik shakllarni aniqlash masalasi ko‘p so‘zli birliklarni (Multiword Expressions – MWE) aniqlash muammosi bilan chambarchas bog‘liq. Analitik fe‘llar ham ko‘p komponentli birliklar sifatida yagona grammatik funksiyani bajaradi.

Avtomatik tanib olishning asosiy bosqichlari quyidagilardan iborat:

1. Matnni tokenlarga ajratish;
2. Morfologik tahlil;
3. Morfologik teglash;
4. Sintaktik bog‘lanishlarni aniqlash;
5. Analitik shakllarni identifikatsiya qilish;
6. Omonim konstruksiyalarni farqlash.

Mazkur bosqichlarning har birida xatolik yuzaga kelishi umumiy aniqlik darajasiga salbiy ta‘sir ko‘rsatadi.

Asosiy muammolar: 1. Omonimiya muammosi. Fe‘l analitik shakllarini avtomatik tanib olishdagi eng murakkab masalalardan biri omonimiyadir.

Masalan: ko‘rib chiqmoq; olib chiqmoq; bo‘lib ber

Shuningdek: yozib qo'ydi; daftarni stolga qo'ydi.

Birinchi holatda *qo'ymoq* yordamchi fe'l vazifasida kelgan bo'lsa, ikkinchi holatda mustaqil leksik ma'noda qo'llangan.

Morfologik analizator faqat so'z shakllariga tayansa, ushbu holatlarni ajratib bera olmaydi. Shuning uchun kontekstual tahlil zarur bo'ladi.

2. Sintaktik va analitik konstruksiyalar o'rtasidagi chegaraning noaniqligi. Ko'plab holatlarda ravishdosh + fe'l ko'rinishidagi birikmalar analitik shakl yoki sintaktik so'z birikmasi bo'lishi mumkin.

Masalan: kulib yubordi; kulib gapirdi.

Birinchi birikmada *yubormoq* yordamchi fe'l bo'lib, kutilmaganlik ma'nosini yuzaga keltiradi. Ikkinchi holatda esa *kulib* ravishdosh, *gapirdi* esa mustaqil kesim hisoblanadi.

Bunday holatlarni avtomatik aniqlash uchun sintaktik bog'lanishlar va semantik munosabatlar tahlili talab etiladi.

3. Yordamchi fe'llarning ko'p funksiyaliligi. O'zbek tilida yordamchi fe'l sifatida ishlatiladigan birliklar asosan mustaqil fe'l sifatida qo'llanadi.

Masalan: turmoq; yurmoq; qolmoq; ketmoq; yubormoq; bermoq; qo'ymoq.

Bu fe'llar ayrim kontekstlarda grammatik ma'no hosil qilsa, boshqa holatlarda mustaqil leksik ma'noda ishlatiladi.

Masalan: ishlay tur; eshik oldida tur.

Birinchi holatda grammatik vazifa, ikkinchisida esa mustaqil harakat ma'nosi mavjud. Natijada avtomatik tizimlar uchun yordamchi va mustaqil qo'llanishlarni ajratish murakkablashadi.

4. Korpus ma'lumotlarining yetarli emasligi. Sun'iy intellekt va mashinaviy o'qitish usullarining muvaffaqiyati katta hajmdagi belgilangan korpuslarga bog'liq.

O'zbek tilida: analitik shakllar korpusi; teglangan morfologik korpus; sintaktik banklar hali yetarli darajada shakllanmagan. Shu sababli mashinaviy o'qitish modellari uchun o'quv ma'lumotlari cheklangan bo'lib qolmoqda.

5. Variantlilik muammosi. Analitik shakllar turli variantlarda uchrashi mumkin. Masalan: aytib yubordi; aytib yuborgan edi; aytib yuborayotgan ekan; aytib yubormoqchi bo'ldi.

Bu shakllarning barchasida asosiy analitik model saqlanib qoladi, biroq tashqi grammatik ko'rsatkichlar o'zgaradi. Qoidalarga asoslangan tizimlar bunday variantlarning hammasini qamrab olishi qiyin.

Avtomatik tanib olish usullari

Qoidalarga asoslangan yondashuv. Bu usulda analitik shakllar maxsus lingvistik qoidalar yordamida aniqlanadi.

Masalan: [Ravishdosh + yordamchi fe'l] modeli bo'yicha quyidagi konstruksiyalar taniladi: aytib berdi; yozib qo'ydi; olib qoldi; kelib qoldi.

Afzalligi: yuqori aniqlik; izohlanadigan natija.

Kamchiligi: qoidalar sonining ortib ketishi; yangi holatlarni qamrab olish qiyinligi.

Statistik usullar. Statistik modellar korpus ma'lumotlari asosida ehtimolliklarni hisoblaydi.

Ular: Hidden Markov Model (HMM); Conditional Random Fields (CRF) kabi algoritmlardan foydalanadi.

CRF modeli qo'shni birliklar va kontekstni hisobga olganligi sababli analitik shakllarni aniqlashda samaraliroq natija beradi.

Neyron tarmoqlar asosidagi yondashuv

So'nggi yillarda Transformer arxitekturasi asosidagi modellar katta muvaffaqiyat qozonmoqda.

Jumladan: BERT; RoBERTa; XLM-R; mBERT.

Mazkur modellar kontekstni chuqur tahlil qilib, yordamchi fe'lining qaysi vazifada qo'llanayotganini aniqlashi mumkin.

Masalan: kitobni stolga qo'ydi; yozib qo'ydi.

BERT modeli kontekstual embeddinglar yordamida ushbu ikki holatni farqlash imkoniyatiga ega.

Disambiguatsiya masalasi. Analitik shakllarni avtomatik tanib olishning markaziy muammosi disambiguatsiya hisoblanadi.

Disambiguatsiya quyidagi omillarga asoslanishi mumkin:

1. Morfologik teglar;
2. Sintaktik bog'lanishlar;
3. Semantik rollar;
4. So'zlarning kollokatsion xususiyatlari;
5. Keng kontekst.

Masalan: aytib berdi; kitob berdi.

Birinchi holatda yordamchi fe'l, ikkinchisida esa mustaqil fe'l mavjud.

Bu farqni aniqlash uchun kontekstual model zarur bo'ladi.

Morfoanalizator yaratishda analitik shakllarning o'рни. O'zbek tili morfoanalizatorida analitik shakllarni alohida birlik sifatida modellashtirish muhim ahamiyatga ega.

Agar *aytib yubordi* konstruktsiyasi ikki mustaqil fe'l sifatida tahlil qilinsa, grammatik ma'no yo'qoladi. Shu sababli morfoanalizator quyidagi ma'lumotlarni qayd etishi lozim: asosiy fe'l; yordamchi fe'l; analitik model turi; grammatik ma'no; zamon; mayl; shaxs-son ko'rsatkichlari.

Bu yondashuv mashina tarjimasini va avtomatik sintaktik tahlil sifatini sezilarli darajada oshiradi. O'zbek tilida fe'l analitik shakllarini avtomatik tanib olish kompyuter lingvistikasining murakkab va dolzarb masalalaridan biridir. Ushbu muammo tilning agglutinatib tabiati, yordamchi fe'llarning ko'p funksiyaliligi, omonimiya, sintaktik va analitik konstruktsiyalar o'rtasidagi chegaralarning noaniqligi hamda belgilangan

korpuslarning yetarli emasligi bilan bog‘liq. Analitik shakllarni muvaffaqiyatli aniqlash uchun qoidalarga asoslangan yondashuvlar, statistik modellar va zamonaviy neyron tarmoqlar imkoniyatlarini birlashtirgan gibrid tizimlardan foydalanish maqsadga muvofiqdir. Ayniqsa, BERT va CRF kabi kontekstual modellarning qo‘llanilishi analitik fe‘llarni boshqa konstruksiyalardan ajratish samaradorligini oshiradi. Kelgusida o‘zbek tili uchun maxsus teglangan korpuslar yaratish, analitik shakllar lug‘atini ishlab chiqish va disambiguatsiya algoritmlarini takomillashtirish ushbu yo‘nalishdagi ilmiy tadqiqotlarning asosiy vazifalaridan biri bo‘lib qoladi.

Xulosa. O‘zbek tilida fe‘l analitik shakllarini avtomatik tanib olish kompyuter lingvistikasi va tabiiy tilni qayta ishlash sohasida murakkab va dolzarb muammolardan biridir. Tadqiqot natijalari shuni ko‘rsatadiki, analitik fe‘l shakllarini aniqlashda omonimiya, yordamchi fe‘llarning ko‘p funksiyaliligi, sintaktik va analitik konstruksiyalar o‘rtasidagi chegaralarning noaniqligi hamda korpus ma‘lumotlarining yetarli emasligi asosiy qiyinchiliklarni tashkil etadi.

Shu bilan birga, qoidalarga asoslangan, statistik va neyron tarmoqlar asosidagi yondashuvlarning uyg‘un qo‘llanilishi analitik shakllarni aniqlash samaradorligini oshiradi. Xususan, BERT va CRF kabi modellar kontekstni chuqur tahlil qilish orqali disambiguatsiya jarayonini yaxshilaydi.

Foydalanilgan adabiyotlar ro‘yxati:

1. Abduazizov A.A. Tilshunoslik nazariyasiga kirish. – Toshkent: Sharq, 2010. – 176 b.
2. Abdullayeva O. O‘zbek tilining internet axborot matnlari korpusini shakllantirishning nazariy va amaliy asoslari: Filol. fan. bo‘yicha falsafa doktori (PhD) diss. – Andijon, 2022. – 120 b.
3. Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида): Филол. фан. бўйича фалсафа доктори (PhD) дисс. автореф. – Тошкент, 2018. – 52 б.
4. Elov B., Hamroyeva Sh., Xusainova Z. NLP (tabiiy tilga ishlov berish)ning vazifalari va zamonaviy yondashuvlar // Filologik tadqiqotlar: til, adabiyot, ta’lim. – Termiz, 2022. – B. 5-6.
5. Hamroyeva Sh., Xusainova Z. NLP (tabiiy tilga ishlov berish)ning vazifalari va zamonaviy yondashuvlar // Filologik tadqiqotlar: til, adabiyot, ta’lim. – Termiz, 2022. – B. 5-6.
6. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.
7. Ҳожиёв А. Тўлиқсиз феъллар. – Тошкент: Фан, 1970. – 263 б.
8. Ҳожиёв А. Ўзбек тилида кўмакчи феъллар. – Тошкент: Фан, 1966. – 223 б.