

O'ZBEK TILINING BAZIS SO'ZLARIGA ASOSLANGAN KONTEKST- ERKLI GRAMMATIKA ISHLAB CHIQISH VA UNING SINTEZI

Nilufar Soliyeva Qahramon qizi,
magistrant
nilufarsoliyeva732@gmail.com
UrDU

Annotatsiya. Mazkur tezisdagi kontekst-erkli grammatika tushunchasining nazariy asoslari, uning tabiiy tilni qayta ishlashdagi ahamiyati va o'zbek tiliga moslashuv imkoniyatlari tahlil qilinadi. Ayniqsa, o'zbek tilining morfologik xususiyatlari, sintaktik qurilishi va bazis (asosiy) so'zlar tarkibini aniqlash mezonlari asosida kontekst-erkli model uchun zarur bo'lgan grammatik tuzilmalarni shakllantirish imkoniyatlari ko'rib chiqiladi. Shuningdek, mavjud grammatik yondashuvlar bilan solishtirma tahlil o'tkazilib, ushbu yondashuvning afzalliklari va qo'llanish muammolari haqida xulosa qilinadi. Ushbu tahliliy yondashuv o'zbek tilini avtomatik qayta ishlash, xususan, mashina tarjimasini, avtomatik tahlil va sintez jarayonlarida yangi ilmiy asos bo'lishi mumkin.

Abstract. This thesis examines the theoretical foundations of context-free grammar, its significance in natural language processing, and its adaptability to the Uzbek language. Special attention is given to the morphological features and syntactic structure of Uzbek, along with the criteria for identifying basic (core) vocabulary, in order to construct the necessary grammatical structures for a context-free model. A comparative analysis is also conducted with existing grammatical approaches, outlining the advantages and limitations of this method. This analytical perspective may serve as a new scientific foundation for automatic processing of the Uzbek language, particularly in machine translation, automatic analysis, and synthesis tasks.

Аннотация. В данной тезисной работе рассматриваются теоретические основы понятия контекстно-свободной грамматики, её значение в обработке естественного языка и возможности адаптации к узбекскому языку. Особое внимание уделяется морфологическим особенностям узбекского языка, его синтаксической структуре, а также критериям определения базисных слов с целью формирования необходимых грамматических конструкций для модели КСГ. Кроме того, проводится сравнительный анализ с существующими грамматическими подходами, выделяются преимущества и проблемы применения данного подхода. Представленный аналитический обзор может служить научной основой для автоматической обработки узбекского языка, в

частности в задачах машинного перевода, автоматического анализа и синтеза.

Kalit so‘zlar: *kontekst-erkli grammatika, o‘zbek tili, sintaksis, morfologiya, bazis so‘zlar, tabiiy tilni qayta ishlash, mashina tarjimasi.*

Zamonaviy sun'iy intellekt va tabiiy tilni qayta ishlash (NLP) texnologiyalarining rivojlanishi natijasida til strukturasi va uning kompyuter tomonidan avtomatik tarzda qayta ishlanishi dolzarb muammoga aylangan. Bu jarayonda tilning sintaktik va morfologik xususiyatlarini formal modellar asosida ifodalash alohida ilmiy yo'nalishga aylangan. Ayniqsa, kontekst-erkli grammatika (KEG) modeli shu kabi ehtiyojlarga samarali yechim sifatida ko'plab tillarda qo'llanilib, muvaffaqiyatli natijalar bermoqda. Ushbu model yordamida grammatik qoidalarga asoslangan holda gaplarni avtomatik sintez qilish, ularni sintaktik jihatdan tahlil qilish va kompyuter dasturlarida real qo'llash imkoniyati yuzaga keladi. Kontekst-erkli grammatika dastlab Noam Chomsky tomonidan 20-asrning ikkinchi yarmida nazariy jihatdan ilgari surilgan bo'lsa-da, amaliyotda, ayniqsa so'nggi yillarda tabiiy tillarni qayta ishlash, mashina tarjimasi, ovoqli yordamchilar, chatbotlar va avtomatik matn yaratish tizimlarida keng qo'llanila boshlandi. Bu modelning afzalligi shundaki, u har bir til birligining tuzilmaviy rolini aniqlab, uni kontekstda joylashuviga ko'ra ifodalash imkonini beradi. Bunda sintez jarayoni grammatik qoidalarga tayanib, tilning mantiqiy va shakliy tuzilmasini aniq aks ettiradi.

Mazkur tezisda aynan shunday yondashuv ilgari suriladi: o'zbek tilining eng ko'p ishlatiladigan, grammatik va semantik jihatdan faol bazis so'zlari tanlab olinadi, ularning morfologik variantlari va sintaktik holatlari KEG qoidalarini asosida modellashtiriladi, so'ngra bu qoidalar asosida avtomatik sintez mexanizmi shakllantiriladi. Bu jarayon orqali o'zbek tilidagi grammatik jihatdan to'g'ri va kontekstga mos gaplarni yaratish imkoniyati yuzaga keladi. Shuningdek, bu tizim mashina tarjimasi, til o'rgatuvchi platformalar va boshqa NLP tizimlariga samarali integratsiya qilinishi mumkin.

O'zbek tili kontekst-erkli grammatika modellarini ishlab chiqish uchun juda qulay va shu bilan birga murakkab til hisoblanadi. Bu, bir tomondan, o'zbek tilining agglutinativ (yopishuvchi) morfologik tuzilishi, ikkinchi tomondan esa sintaktik erkinlik darajasining nisbatan yuqoriligi bilan bog'liq. O'zbek tilida so'zlar o'zaro qo'shib, grammatik ma'no hosil qiluvchi affikslar yordamida murakkab birliklar tashkil etadi. Shu sababli, har bir birlikning grammatik vazifasi kontekstdan kelib chiqib o'zgaradi. Masalan, “kelgan” so'zi turli kontekstlarda sifatlash, fe'l, yoki hatto gapda kesim vazifasida ishtirok etishi mumkin.

Kontekst-erkli grammatika aynan mana shu o'zgaruvchanlikni tahlil qilish va ifodalashda muhim rol o'ynaydi. Bu modelda har bir qoidaning amal qilishi uchun uning chap va o'ng konteksti aniqlanadi. Bu, o'z navbatida, o'zbek tilida fe'l

shakllarining zamon, nisbat, shaxsga qarab moslashuvi, holatlar bilan bog'lanishi va gapda vazifasining farqlanishini to'g'ri tahlil qilish imkonini beradi. Masalan:

Men **bordim** (fe'l – o'tgan zamon, 1-shaxs, birlik)

Sen **bording** (fe'l – o'tgan zamon, 2-shaxs, birlik)

U **bordi** (fe'l – o'tgan zamon, 3-shaxs, birlik)

Ushbu uchta shakl morfologik jihatdan o'xshash bo'lsa-da, kontekst-erkli model ularning o'zaro farqini tahlil qilish va sintetik yo'l bilan hosil qilishda muhim ahamiyat kasb etadi. O'zbek tilining bazis (asosiy) so'zlariga asoslangan model yaratishda dastlab, tilning eng ko'p ishlatiladigan va grammatik jihatdan eng faol birliklarini aniqlash zarur. Bunday so'zlar, odatda, o'zbek tilining ommaviy matnlar korpusi asosida, chastotali tahlil orqali aniqlanadi. Masalan, “bor”, “kel”, “ol”, “qil”, “bilan”, “uchun”, “men”, “sen”, “u”, “shu”, “bu” kabi birliklar bazis korpusining asosini tashkil qilishi mumkin. Bazis so'zlarga asoslangan kontekst-erkli grammatika yaratishda yana bir muhim bosqich – bu bazis birliklarning sintaktik va morfologik variantlarini tizimlashtirishdir. Har bir so'z o'zining imkonli kombinatorik holatlari, ya'ni qanday affikslar bilan qanday kontekstlarda bog'lanishi mumkinligi bo'yicha tahlil qilinadi. Bu jarayonda qoidaga asoslangan (rule-based) va statistik usullar qo'llanishi mumkin. Masalan, “ol-” fe'li:

oldi – o'tgan zamon, 3-shaxs

oldim – o'tgan zamon, 1-shaxs

olgan – sifatdash

olindi – majhul nisbatda

Shuningdek, bu shakllarning kontekstda qanday ishlatilishini aniqlashda avtomatik tahlil tizimlarining yordami zarur bo'ladi. Bu yerda mashina o'rganishi (machine learning) yondashuvlari, ayniqsa neyron tarmoqlar asosida ishlovchi modellar katta ahamiyat kasb etadi. Ammo ularga kontekst-erkli qoida tizimi integratsiya qilinmaguncha, bunday modellar o'zbek tilining murakkab grammatikasini to'liq o'zlashtira olmaydi. Bundan tashqari, sintetik grammatika — ya'ni, til birliklarini yaratish (sintez) yo'li bilan yangi shakllarni hosil qilish — kontekst-erkli yondashuv asosida ancha ishonchli va moslashuvchan bo'ladi. Masalan, avtomatik matn yaratish yoki mashina tarjimai tizimida o'zbek tilining grammatik shakllari foydalanuvchining kontekstiga mos tarzda sintez qilinadi: ya'ni, model “sen uchun yozgan edim” shaklini foydalanuvchi ifodalayotgan nuqtai nazarga mos shaklda yaratadi. O'zbek tilida erkin so'z tartibi mavjudligi (masalan, “men kitobni o'qidim”, “kitobni men o'qidim”, “o'qidim men kitobni”) kontekst-erkli yondashuvni talab qiladi. So'z tartibining erkinligi formal qoidalarga

asoslangan modelda chalkashliklarga olib keladi, kontekst-erkli model esa bu holatlarni mantiqiy kontekst orqali farqlashga qodir.

Tabiiy tilni avtomatik tarzda qayta ishlash texnologiyalari hozirgi kunda sun'iy intellekt sohasida katta rol o'ynamoqda. Bu texnologiyalarning muhim yo'nalishlaridan biri — til sintezi bo'lib, u grammatik jihatdan to'g'ri gaplar va matnlar yaratish imkonini beradi. Sintez jarayonida ishlatiladigan asosiy model – bu formal grammatikalardan biri bo'lgan kontekst-erkli grammatika (KEG) hisoblanadi. KEG yordamida tilning sintaktik tuzilmasi qat'iy qoidalar asosida tavsiflanadi va bu qoidalar asosida yangi gaplarni yaratish imkoniyati tug'iladi. KEG modeli to'rtta asosiy komponentdan tashkil topadi:

N – no-terminal belgilar to'plami, Σ – terminal belgilar (ya'ni real so'zlar), P – produktsiyalar (o'zgartirish qoidalari), S – boshlang'ich belgi (odatda, gapni anglatadi). KEG qoidalari yordamida til strukturasi daraxtsimon shaklda ifodalanadi va sintaktik bog'liqliklar aniqlanadi.

Sintez jarayoni KEG modelidagi qoidalarining teskari yo'nalishda ishlatilishi orqali amalga oshiriladi. Ya'ni, avval boshlang'ich belgi — odatda S belgisi olinadi, so'ngra unga mos produktsiya qoidasi tanlanadi va u bosqichma-bosqich terminal belgilar to'plamiga aylantiriladi. Natijada, grammatik jihatdan to'g'ri gap yoki ibora hosil qilinadi. Bu jarayon, kompyuter dasturlari yordamida avtomatlashtirilgan holda bajariladi va real til sintezi amaliyotida juda muhim ahamiyat kasb etadi.

Misol tariqasida quyidagi qoidalarni olaylik:

$S \rightarrow NP VP$
 $NP \rightarrow Det N$
 $VP \rightarrow V NP$
 $Det \rightarrow bu | bir$
 $N \rightarrow bola | o'quvchi | kitob$
 $V \rightarrow o'qiydi | yozadi | oladi$

Ushbu qoidalardan foydalangan holda sintez jarayoni quyidagicha kechadi:

- 1) Boshlang'ich belgi sifatida S belgilanadi.
- 2) $S \rightarrow NP VP$ qoidasi asosida NP va VP hosil qilinadi.
- 3) $NP \rightarrow Det N$ qoidasi orqali NP belgisi “bu bola” shaklida sintez qilinadi.
- 4) $VP \rightarrow V NP$ esa “o'qiydi kitob” kabi segmentlarni beradi. Yakuniy natijada, masalan, “Bu bola kitob o'qiydi” gapini hosil qilish mumkin bo'ladi. Bu esa til sintezining eng sodda, lekin amaliy modelidir.

Yuqoridagi jarayon davomida hosil bo'lgan natijani sintaktik daraxt (parse tree) shaklida ham ifodalash mumkin. Bu daraxt tildagi har bir birikmaning qanday

grammatik komponentlardan tashkil topganini ko'rsatadi. Masalan, S – gap elementi NP (ot birikmasi) va VP (fe'l birikmasi) orqali tuzilganini, NP esa Det (aniqllovchi) va N (ot) orqali ifodalanganini ochiq-oydin ko'rsatadi. Bu sintaktik daraxt KEG asosidagi sintez mexanizmining vizual va tushunarli ifodasidir.

Shuningdek, KEG asosidagi sintez mexanizmi til strukturasi qat'iy bo'lgan agglutinatив tillar, jumladan o'zbek tili uchun ham moslashtirilishi mumkin. O'zbek tilining morfologik jihatdan boyligi (ya'ni qo'shimchalar tizimi), SOV tartibi (ega-to'ldiruvchi-kesim) va kesimning gap oxirida kelishi KEG qoidalari orqali ifodalanib, sintez mexanizmida muvaffaqiyatli qo'llanishi mumkin. Bunda bazis so'zlar (eng ko'p uchraydigan va grammatik jihatdan markaziy so'zlar) KEGning terminal belgilariga aylanadi. No-terminal belgilar esa tilning asosiy sintaktik birliklarini tashkil qiladi: masalan, NP, VP, PP (predlogli birikma) va hokazo.

Sintez mexanizmining algoritmik bosqichlari quyidagicha ifodalanadi:

1. Boshlang'ich belgi tanlanadi (S);
2. S ga mos produktsiya qoidalari tanlanadi;
3. Har bir no-terminal belgi mos terminal yoki no-terminal birliklarga aylantiriladi;
4. Terminal belgilar orqali gap yakuniy ko'rinishda hosil qilinadi.

Ushbu ketma-ketlik asosida KEG sintez mexanizmi hatto murakkab gaplarni hosil qilishga imkon beradi. Masalan, ergash gaplar, bog'lovchili yoki bog'lovchisiz gaplar, holli fe'l shakllari, inkor shakllari ham ushbu model doirasida ifodalanishi mumkin. Ammo KEG asosidagi sintez modelida ayrim cheklovlar ham mavjud. Ular orasida semantik nazoratning cheklanganligi, stilistik moslikni saqlab qolishdagi qiyinchiliklar va morfologik variantlarning ko'pligi tufayli qoidalar sonining ko'payib ketishi kabi holatlar mavjud. Biroq zamonaviy NLP tizimlarida KEGni semantik tarmoqlar, lemmatizatorlar va tahlilchilar bilan birgalikda ishlatish orqali bu muammolarni kamaytirish mumkin. Kontekst-erkli grammatika (KEG) modeli zamonaviy tabiiy tilni qayta ishlash (NLP) texnologiyalarida muhim o'rin tutadi. U tilning sintaktik va morfologik tuzilishini qat'iy qoidalarga asoslangan holda modellashtirish, tahlil qilish va sintez qilish imkoniyatini beradi. KEG ayniqsa grammatik jihatdan murakkab va agglutinatив tillar, jumladan o'zbek tili uchun ham nihoyatda foydali model hisoblanadi. Ushbu tadqiqotda KEG modeli asosida o'zbek tilining bazis so'zlariga tayangan holda sintez mexanizmini shakllantirish masalasi atroflicha o'rganildi.

Tahlillar shuni ko'rsatdiki, o'zbek tilining so'z yasovchanligi, affikslar tizimi va erkin so'z tartibi KEG modelini qo'llashda alohida yondashuvni talab etadi. Shu sababli, bazis (eng ko'p ishlatiladigan) so'zlarni aniqlash, ularning grammatik variantlarini sistemalashtirish va KEG qoidalari orqali sintez algoritmini ishlab

chiqish zarur bo'ldi. Tadqiqotda bu bosqichlar izchil tarzda amalga oshirildi: no-terminal va terminal belgilar aniqlandi, produktsiya qoidalari ishlab chiqildi va real gaplar sintez qilish imkoniyati sinovdan o'tkazildi.

KEG asosidagi sintez mexanizmi orqali o'zbek tilida grammatik jihatdan to'g'ri, kontekstga mos va mantiqiy gaplar yaratish mumkinligi amalda isbotlandi. Bunday yondashuv mashina tarjimasi, avtomatik matn sintezi, chatbotlar va ovozli yordamchilar kabi zamonaviy texnologiyalar uchun poydevor vazifasini bajaradi. Shuningdek, tilni o'rgatish, grammatik strukturani tushuntirish kabi ta'limiy maqsadlar uchun ham KEG modeli asosidagi yondashuvlar samarali qo'llanishi mumkin. Shuni alohida ta'kidlash lozimki, KEG modeli o'zbek tilining formal qoidalarini chuqur o'rganish, sintaktik va morfologik birliklarni tizimli shaklda tasniflash, ularni algoritmik tarzda tahlil qilish va avtomatik tarzda gap yaratish jarayonida ilmiy asos bo'lib xizmat qiladi. Kelgusida bu yondashuvni semantik tahlil tizimlari, neyron tarmoqlarga asoslangan modellar, korpus lingvistikasi va leksikografik tadqiqotlar bilan integratsiyalash orqali yanada mukammallashtirish mumkin.

Foydalanilgan adabiyotlar:

1. N. Chomsky, Syntactic Structures, 2nd ed. Berlin: Mouton de Gruyter, 2002.
2. D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. Draft, Stanford University, 2023. [Online]. Available: <https://web.stanford.edu/~jurafsky/slp3/>
3. A. V. Aho and J. D. Ullman, The Theory of Parsing, Translation, and Compiling: Volume I: Parsing. Englewood Cliffs, NJ: Prentice-Hall, 1972.
4. K. Knight, "Automated Knowledge Acquisition for Machine Translation," Machine Translation, vol. 5, no. 2, pp. 129–144, 1990.
5. D. Klein and C. D. Manning, "Accurate unlexicalized parsing," in Proc. 41st Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan, 2003, pp. 423–430.