



O‘ZBEK TILIDA SINTAKTIK ANALIZATOR YARATISH TAJRIBASI (QOIDAVIY MODEL ASOSIDA)

*Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili
va adabiyoti universiteti 3-bosqich talabasi*
Sayfiddinova Ziynat Abduzohidovna

Annotatsiya: Mazkur maqolada o‘zbek tilida avtomatik sintaktik tahlilni amalga oshirishga qaratilgan qoidaviy (rule-based) sintaktik analizator yaratish tajribasi yoritiladi. Tadqiqot doirasida o‘zbek tilining morfologik va sintaktik xususiyatlariga asoslangan sodda analizator modeli ishlab chiqildi hamda u tanlab olingan gaplar to‘plami asosida sinovdan o‘tkazildi. Analizatorning ishlash jarayonida kelishik ko‘rsatkichlari, so‘z turkumlari hamda boshqaruv va moslashuv munosabatlari asosiy mezon sifatida belgilandi. Olingan natijalar tahlili analizatorning ma‘lum turdagi sodda gaplarda nisbatan yuqori aniqlikda ishlashini, murakkab sintaktik tuzilmalarda esa xatoliklar kuzatilishini ko‘rsatdi. Tadqiqot yakunida aniqlangan muammolar asosida sintaktik analizatorni takomillashtirishning istiqbolli yo‘nalishlari bo‘yicha tavsiyalar berildi. Ushbu ish o‘zbek tili uchun avtomatik sintaktik tahlil tizimlarini ishlab chiqish bo‘yicha dastlabki amaliy tajriba sifatida ahamiyatga ega.

Kalit so‘zlar: o‘zbek tili, sintaktik analizator, avtomatik gap tahlili, qoidaviy model, boshqaruv munosabati.

Kirish

So‘nggi yillarda tabiiy tillarni avtomatik qayta ishlash (Natural Language Processing) sohasida olib borilayotgan tadqiqotlar jadal rivojlanib, turli tillar uchun morfologik va sintaktik tahlilga mo‘ljallangan tizimlar yaratildi. Biroq agglutinativ tillar sirasiga mansub bo‘lgan o‘zbek tili uchun avtomatik sintaktik tahlil masalasi hanuz dolzarb va murakkab muammolardan biri bo‘lib qolmoqda. Bunga o‘zbek tilining boy morfologik tizimi, kelishik shakllarining ko‘p ma‘noliligi hamda sintaktik munosabatlarning turli shakllarda ifodalanishi sabab bo‘lmoqda.

Sintaktik analiz til birliklari o‘rtasidagi munosabatlarni aniqlashda muhim bosqich bo‘lib, gap bo‘laklari, ularning o‘zaro bog‘lanishi hamda boshqaruv va moslashuv munosabatlarini aniqlashni nazarda tutadi. An‘anaviy tilshunoslikda bu masalalar keng va chuqur o‘rganilgan bo‘lsa-da, ushbu bilimlarni kompyuter tomonidan avtomatik aniqlash imkonini beruvchi aniq va formal qoidalarga aylantirish muammosi yetarli darajada hal etilmagan. Ayniqsa, o‘zbek tilida kelishik ko‘rsatkichlari orqali ifodalanadigan sintaktik munosabatlarni avtomatik aniqlash alohida e‘tibor talab etadi.



Mavjud tadqiqotlarda statistik va neyron yondashuvlar tobora keng qo'llanilayotgan bo'lsa-da, resurslarning cheklanganligi sharoitida qoidaviy (rule-based) yondashuv o'z ahamiyatini saqlab qolmoqda. Qoidaviy analizatorlar tilning grammatik qonuniyatlariga bevosita tayanishi, tahlil jarayonining shaffofligi va izohlanishi bilan ajralib turadi. Shu sababli dastlabki bosqichda sintaktik tahlil tizimini yaratishda qoidaviy model asosidagi yechimlar muhim ilmiy va amaliy ahamiyat kasb etadi.

Mazkur maqolada o'zbek tili uchun qoidaviy yondashuvga asoslangan sodda sintaktik analizator yaratish tajribasi yoritiladi. Tadqiqotning asosiy maqsadi — ishlab chiqilgan analizatorning imkoniyatlarini sinovdan o'tkazish, uning aniqlik darajasini baholash hamda yuzaga keladigan xatolarni tahlil qilishdan iborat. Shuningdek, olingan natijalar asosida o'zbek tilida avtomatik sintaktik tahlil tizimlarini yanada takomillashtirish yo'nalishlari bo'yicha tavsiyalar ishlab chiqiladi. Mazkur tadqiqot quyidagi savolga javob berishni maqsad qiladi: qoidaviy yondashuv o'zbek tilidagi sodda gaplarni avtomatik sintaktik tahlil qilishda qanchalik samarali bo'la oladi?

Metodologiya

Mazkur tadqiqotda o'zbek tilida avtomatik sintaktik tahlilni amalga oshirish uchun qoidaviy (rule-based) yondashuvga asoslangan sodda sintaktik analizator modeli ishlab chiqildi. Tadqiqot metodologiyasi an'anaviy o'zbek tilshunosligida shakllangan sintaktik qoidalarni kompyuter tomonidan qayta ishlash imkonini beruvchi formal ko'rinishga keltirishga asoslanadi. Analizator prototip ko'rinishida bo'lib, uning asosiy vazifasi gap tarkibidagi so'zlar o'rtasidagi sintaktik munosabatlarni aniqlashdan iborat.

Tadqiqot materiallari sifatida hozirgi o'zbek adabiy tiliga mansub bo'lgan sodda va yoyiq gaplardan tashkil topgan cheklangan hajmdagi gaplar to'plami tanlandi. Sinov uchun jami 100 ta gap qo'llanildi. Gaplar asosan badiiy va ilmiy-uslubiy matnlardan tanlab olindi hamda ularning sintaktik tuzilishi murakkab bo'lmagan, ya'ni qo'shma va ergash gapli konstruktsiyalarni o'z ichiga olmaydigan birliklardan iborat ekanligi e'tiborga olindi. Bu cheklov analizatorning dastlabki bosqichdagi imkoniyatlarini aniq baholash maqsadida joriy etildi. Sinovdan oldin barcha gaplar qo'lda sintaktik tahlil qilinib, etalon (to'g'ri) tahlil natijalari shakllantirildi. Keyinchalik avtomatik tahlil natijalari aynan shu etalon bilan solishtirildi.

Analizator modeli va yondashuv

Ishlab chiqilgan sintaktik analizator qoidaviy model asosida faoliyat yuritadi. Ushbu modelda sintaktik munosabatlarni aniqlashda kelishik ko'rsatkichlari, so'z turkumlari hamda boshqaruv va moslashuv munosabatlari asosiy mezon sifatida belgilandi. Analizator statistik yoki mashinaviy o'rganish usullaridan foydalanmaydi, balki oldindan belgilangan grammatik qoidalarga



tayanadi.

Sintaktik tahlil quyidagi asosiy bosqichlarda amalga oshirildi:

Tokenlash bosqichi. Kiruvchi gap so'zlarga ajratildi va har bir so'z alohida birlik sifatida qayd etildi.

So'z turkumini aniqlash. Har bir token uchun oldindan tuzilgan leksik ro'yxat va grammatik belgilarga asoslangan holda so'z turkumi belgilandi (ot, fe'l, sifat, son va boshqalar).

Morfologik ko'rsatkichlarni aniqlash. So'z tarkibidagi kelishik, egalik va fe'l shakllari aniqlanib, ularning sintaktik vazifaga ta'siri hisobga olindi.

Sintaktik munosabatlarni belgilash. Aniqlangan morfologik ko'rsatkichlar asosida gap bo'laklari o'rtasidagi boshqaruv va moslashuv munosabatlari aniqlanib, ularning qaysi gap bo'lagiga tegishli ekanligi belgilandi.

Mazkur bosqichlar ketma-ketligi analizatorning ichki ishlash mexanizmini tashkil etadi va har bir bosqich keyingi bosqich uchun tayanch vazifasini bajaradi.

Baholash mezonlari

Analizatorning ishlash samaradorligini baholash uchun aniqlik ko'rsatkichi (accuracy) asosiy mezon sifatida tanlandi. Har bir gap uchun avtomatik tahlil natijalari qo'lda tayyorlangan etalon tahlil bilan solishtirildi. To'g'ri aniqlangan sintaktik bog'lanishlar sonining umumiy bog'lanishlar soniga nisbati foiz ko'rinishida hisoblandi.

Baholash jarayonida quyidagi jihatlar alohida e'tiborga olindi:

- gap bo'laklarining to'g'ri aniqlanishi;
- kelishik ko'rsatkichlarining sintaktik vazifaga mos talqin qilinishi;
- boshqaruv va moslashuv munosabatlarining aniqligi.

Metodologik cheklovlar

Tadqiqot metodologiyasi doirasida bir qator cheklovlar mavjud. Xususan, analizator faqat sodda gaplarni qamrab oladi va murakkab sintaktik konstruktsiyalarni to'liq tahlil qilish imkoniyatiga ega emas. Shuningdek, kelishiklarning ko'p ma'noliligi va kontekstga bog'liq talqinlari barcha holatlarda qoidaviy yondashuv orqali to'liq aniqlanmaydi. Ushbu cheklovlar tadqiqot natijalarini talqin qilishda inobatga olindi hamda keyingi bosqichlar uchun yo'nalish sifatida belgilandi.

Natijalar

Tadqiqot doirasida ishlab chiqilgan qoidaviy sintaktik analizator tanlab olingan gaplar to'plamida sinovdan o'tkazildi. Sinov jarayonida analizatorning sintaktik munosabatlarni aniqlashdagi umumiy samaradorligi hamda alohida



komponentlar bo'yicha ishlash natijalari baholandi. Natijalar qo'lda tayyorlangan etalon tahlil bilan solishtirish asosida olindi.

Umumiy aniqlik ko'rsatkichlari

Sinov uchun tanlangan 100 ta gap avtomatik sintaktik tahlildan o'tkazildi. Tahlil natijalariga ko'ra, analizator tomonidan aniqlangan sintaktik bog'lanishlarning ma'lum qismi etalon tahlilga mos keldi. Umumiy hisobda analizatorning ishlash aniqligi 74% oralig'ida ekanligi kuzatildi. Bu ko'rsatkich analizatorning sodda va morfologik jihatdan aniq shakllangan gaplarda nisbatan barqaror ishlashini ko'rsatadi.

Gap bo'laklarini aniqlash bo'yicha natijalar shuni ko'rsatdiki, ega va kesim munosabatlari aksariyat hollarda to'g'ri aniqlangan. Ayniqsa, fe'l kesim bilan kelgan sodda gaplarda analizatorning aniqlik darajasi yuqori bo'ldi. Ikkinchi darajali gap bo'laklari, xususan, to'ldiruvchi va aniqlovchini aniqlashda esa ayrim xatoliklar qayd etildi.

Kelishiklar asosida tahlil natijalari

Kelishik ko'rsatkichlari orqali ifodalanadigan sintaktik munosabatlarni aniqlash natijalari alohida tahlil qilindi. Tushum va jo'nalish kelishiklarida ifodalangan to'ldiruvchilar ko'p hollarda to'g'ri aniqlangan bo'lsa-da, chiqish va o'rin-payt kelishiklarida ko'p ma'nolilik sababli xatoliklar ko'proq uchradi. Ayniqsa, o'rin-payt kelishigi bilan ifodalangan hol bo'laklari ba'zi hollarda to'ldiruvchi sifatida noto'g'ri talqin qilindi.

Qaratqich kelishigi orqali ifodalangan aniqlovchilarni aniqlashda analizator nisbatan barqaror natijalar ko'rsatdi. Biroq egalik qo'shimchasi bilan kelgan, lekin qaratqich munosabatini talab qilmaydigan birikmalarda xatoliklar qayd etildi. Bu holat kelishik shaklining formal belgilariga ortiqcha tayanish bilan izohlanadi.

Sintaktik munosabatlar bo'yicha natijalar

Boshqaruv va moslashuv munosabatlarini aniqlash natijalari ham tahlil qilindi. Moslashuv munosabatlari, xususan, ega-kesim orasidagi shaxs-son moslashuvi, aksariyat hollarda to'g'ri belgilandi. Boshqaruv munosabatlarida esa kelishiklarning semantik jihatdan turlicha talqin qilinishi analizator ishiga salbiy ta'sir ko'rsatdi.

Ba'zi gaplarda bir so'zning bir nechta sintaktik vazifani bajarish ehtimoli mavjud bo'lib, bunday holatlarda analizator faqat bitta qoidaviy variantni tanladi. Natijada alternativ tahlil imkoniyati cheklanganligi sababli noto'g'ri natijalar yuzaga keldi.

Xatoliklar tahlili

Natijalar shuni ko'rsatdiki, analizatorida kuzatilgan asosiy xatoliklar quyidagi holatlar bilan bog'liq:



- kelishik shakllarining ko‘p ma’noiligi;
- kontekstga bog‘liq sintaktik vazifalarni aniqlashdagi qiyinchiliklar;
- murakkabroq sintaktik bog‘lanishlarning qoidaviy modelda to‘liq qamrab olinmaganligi.

Shu bilan birga, analizatorning dastlabki model sifatida sodda gaplarni avtomatik sintaktik tahlil qilish imkoniyatiga ega ekanligi tajriba natijalari orqali tasdiqlandi.

Muhokama

Olingan natijalar qoidaviy yondashuvga asoslangan sintaktik analizatorning o‘zbek tilidagi sodda gaplarni avtomatik tahlil qilishda muayyan darajada samarali ekanligini ko‘rsatdi. Umumiy aniqlik ko‘rsatkichining 70% dan oshiq bo‘lishi analizatorning dastlabki prototip sifatida o‘z vazifasini qisman bajarayotganini tasdiqlaydi. Bu holat, avvalo, o‘zbek tilida sintaktik munosabatlarning ko‘plab hollarda morfologik ko‘rsatkichlar orqali ifodalanishi bilan izohlanadi.

Natijalardan ko‘rinib turibdiki, ega–kesim munosabatlarini aniqlashda analizator nisbatan barqaror ishlagan. Buning sababi shundaki, ushbu munosabat o‘zbek tilida ko‘pincha shaxs-son ko‘rsatkichlari orqali formal tarzda ifodalanadi va qoidaviy yondashuv uchun qulay hisoblanadi. Xuddi shuningdek, tushum va jo‘nalish kelishiklarida ifodalangan to‘ldiruvchilarni aniqlashda ham analizator yuqoriroq aniqlik ko‘rsatdi. Bu kelishiklarning sintaktik vazifasi nisbatan aniq va barqarorligi bilan bog‘liq.

Biroq kelishiklarning ko‘p ma’noiligi analizator ishida asosiy muammolardan biri sifatida namoyon bo‘ldi. Xususan, o‘rin-payt va chiqish kelishiklarida ifodalangan hol va to‘ldiruvchi munosabatlarini farqlashda sezilarli xatoliklar kuzatildi. Bu holat qoidaviy modelda kontekstual omillar yetarli darajada hisobga olinmaganligini ko‘rsatadi. An’anaviy tilshunoslikda bu kabi holatlar ko‘pincha semantik va pragmatik mezonlar orqali izohlanadi, biroq ularni formal qoidalarga to‘liq aylantirish murakkab vazifa hisoblanadi.

Qaratqich kelishigi bilan bog‘liq xatoliklar ham muhokamaga loyiqdir. Analizator egalik qo‘shimchasi mavjud bo‘lgan barcha birikmalarni avtomatik tarzda qaratqich munosabatiga ega deb talqin qilgan hollarda noto‘g‘ri natijalar yuzaga keldi. Bu holat egalik va qaratqich munosabatlarining formal jihatdan o‘xshashligi, ammo sintaktik funksiyasi jihatidan farqlanishi bilan izohlanadi. Mazkur muammo qoidaviy yondashuv doirasida qo‘shimcha cheklovlar va istisno qoidalarni ishlab chiqishni talab etadi.

Boshqaruv munosabatlarini aniqlashda kuzatilgan xatoliklar ham tahlil



qilinishi lozim. Ayrim hollarda bir sintaktik birlik bir nechta mumkin bo'lgan munosabatga ega bo'lishi mumkin, biroq ishlab chiqilgan analizator faqat bitta variantni tanlashga majbur bo'ldi. Bu esa alternativ tahlil imkoniyatining cheklanganligini ko'rsatadi. Statistik yoki neyron modellarda ehtimollik asosida bir nechta variantni ko'rib chiqish imkoniyati mavjud bo'lsa-da, qoidaviy yondashuvda bunday moslashuvchanlikni ta'minlash murakkabroqdir.

Shu bilan birga, tadqiqot natijalari qoidaviy yondashuvning muhim afzalliklarini ham namoyon etdi. Xususan, analiz jarayonining shaffofligi, har bir qarorning aniq grammatik qoida bilan izohlanishi va xatolik sabablarini aniqlash imkoniyati ushbu modelning kuchli tomonlari hisoblanadi. Bu jihat qoidaviy analizatorni o'quv, ilmiy va tajriba-sinov maqsadlarida qo'llash uchun qulay qiladi.

Mazkur tadqiqot natijalari mavjud ilmiy ishlarda qayd etilgan xulosalar bilan hamohangdir. Xususan, agglutinatив tillar uchun avtomatik sintaktik tahlilda morfologik boylik bir vaqtning o'zida ham imkoniyat, ham murakkablik manbai ekani avvalgi tadqiqotlarda ham ta'kidlangan. Ushbu ishda olingan natijalar ham aynan shu holatni o'zbek tili misolida yana bir bor tasdiqlaydi.

Umuman olganda, muhokama natijalari shuni ko'rsatadiki, ishlab chiqilgan sintaktik analizator mukammal tizim bo'lmasa-da, o'zbek tilida avtomatik sintaktik tahlilni amalga oshirish yo'lidagi muhim dastlabki qadam hisoblanadi. Aniqlangan muammolar va xatoliklar keyingi bosqichlarda modelni takomillashtirish uchun asos bo'lib xizmat qiladi hamda qoidaviy yondashuvni statistik yoki mashinaviy o'rganish usullari bilan integratsiya qilish zaruratini ko'rsatadi.

Xulosa

Mazkur tadqiqotda o'zbek tili uchun qoidaviy yondashuvga asoslangan sodda sintaktik analizator yaratish tajribasi amalga oshirildi va uning dastlabki ishlash samaradorligi baholandi. Olingan natijalar ishlab chiqilgan analizatorning sodda sintaktik tuzilishga ega gaplarni avtomatik tahlil qilishda ma'lum darajada muvaffaqiyatli ekanligini ko'rsatdi hamda o'zbek tilida sintaktik munosabatlarni avtomatik aniqlash imkoniyatlari mavjudligini tasdiqladi.

Tadqiqot davomida aniqlanganki, morfologik ko'rsatkichlar, xususan kelishik shakllari va shaxs-son moslashuvi sintaktik tahlilda muhim tayanch omil bo'lib xizmat qiladi. Shu bilan birga, kelishiklarning ko'p ma'noliligi, kontekstga bog'liq talqinlar hamda murakkab sintaktik bog'lanishlar qoidaviy model imkoniyatlarini cheklovchi asosiy omillar sifatida namoyon bo'ldi. Bu holat o'zbek tilining lingvistik xususiyatlarini avtomatik tahlilga moslashtirishda yanada chuqurroq yondashuv zarurligini ko'rsatadi.



Ishlab chiqilgan sintaktik analizator mukammal tizim bo'lmasa-da, u o'zbek tilida avtomatik sintaktik tahlil tizimlarini yaratish yo'lidagi dastlabki amaliy qadam sifatida ahamiyatga ega. Tadqiqot natijalari qoidaviy yondashuvning kuchli va zaif tomonlarini aniqlashga imkon berdi hamda kelgusida ushbu yondashuvni statistik va mashinaviy o'rganish usullari bilan integratsiya qilish istiqbollari belgilab berdi. Mazkur tadqiqot natijalari o'zbek tili uchun avtomatik sintaktik tahlil tizimlarini yaratishda qoidaviy yondashuvning dastlabki, ammo zarur bosqich ekanligini ko'rsatadi.

Kelgusidagi tadqiqotlar doirasida analizator qoidalarini kengaytirish, murakkab gap turlarini qamrab olish, katta hajmdagi matnlar korpusida sinovdan o'tkazish hamda ehtimollik asosidagi modellardan foydalanish maqsadga muvofiqdir. Ushbu yo'nalishlar o'zbek tili uchun avtomatik sintaktik tahlil tizimlarining aniqligi va amaliy qo'llanish doirasini kengaytirishga xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR

1. Abdurahmonov G., Shukurov Sh. O'zbek tilining grammatikasi. – Toshkent: Fan, 1975.
2. Hojiyev A. O'zbek tili sintaksisi. – Toshkent: Fan, 1991.
3. Rahmatullayev Sh. Hozirgi o'zbek adabiy tili. – Toshkent: Universitet, 2006.
5. Sayfullayeva R., Mengliyev B. Hozirgi o'zbek adabiy tili. Sintaksis. – Toshkent: Fan va texnologiya, 2010.
6. Jurafsky D., Martin J. H. Speech and Language Processing. – Pearson Education, 2023.
7. Manning C. D., Schutze H. Foundations of Statistical Natural Language Processing. – MIT Press, 1999.
8. Kubler S., McDonald R., Nivre J. Dependency Parsing. – Morgan & Claypool Publishers, 2009.
9. Usmonov I. O'zbek tilshunosligida kompyuter texnologiyalarining qo'llanilishi. – Toshkent: Akademnashr, 2018.