

O‘ZBEK TILI MORFOLOGIK TAHLILIDA KONTEKSTUAL DISAMBIGUATSIYA JARAYONI XUSUSIDA

Barno Xabibulla qizi ISLOMOVA

assistent

Alisher Navoiy nomidagi

O‘zbek tili va adabiyoti universiteti

Toshkent, O‘zbekiston

Annotasiya

Maqolada o‘zbek tili morfologik tahlilida kontekstual disambiguatsiya jarayonining mohiyati, vazifalari va zamonaviy lingvistik texnologiyalardagi roli yoritilgan. Shuningdek, disambiguatsiya jarayonining turlari – leksik, morfologik va sintaktik darajalarda misollar asosida tahlil qilingan hamda sun’iy intellekt modellarida (BERT, GPT, Word2Vec) kontekstual ma’no aniqlash usullari va mazkur yo‘nalishdagi muammolar o‘rganilgan.

Tayanch so‘zlar: kontekstual disambiguatsiya, morfologik tahlil, kontekst, sun’iy intellekt, leksik noaniqlik (ambiguitet), o‘zbek tili, Natural Language Processing (NLP).

О ПРОЦЕССЕ КОНТЕКСТУАЛЬНОЙ ДИСАМБИГУАЦИИ В МОРФОЛОГИЧЕСКОМ АНАЛИЗЕ УЗБЕКСКОГО ЯЗЫКА

Барно Хабибулла кизи ИСЛАМОВА

ассистент

Университет узбекского языка и литературы

имени Алишера Навои

Ташкент, Узбекистан

Аннотация

В статье освещены сущность, задачи и роль в процессе разрешения контекстуальной неоднозначности при морфологическом анализе узбекского языка и современные лингвистические технологии. Также на основе примеров анализируются типы процесса разрешения контекстуальной неоднозначности — на лексическом, морфологическом и синтаксическом уровнях, а также методы определения контекстуального значения и модели искусственного интеллекта (Bert, GPT, Word2Vec), а также проблемы и способы их решения.

Ключевые слова: разрешение контекстная неоднозначность, морфологический анализ, контекст, искусственный интеллект, лексическая неоднозначность, узбекский язык, НЛП.

Hozirgi davrda tabiiy tilni qayta ishlash (NLP – *Natural Language Processing*) texnologiyalarida disambiguatsiya jarayoni eng murakkab, ammo zaruriy bosqichlardan biri hisoblanadi. Tilning tabiiy xususiyatlaridan biri shundaki, bir so‘z bir nechta grammatik shakl yoki ma’noga ega bo‘lishi mumkin.

Shu sababli, har qanday avtomatik tahlil tizimi kontekst asosida to‘g‘ri ma‘noni tanlashni o‘rganishi lozim. O‘zbek tili agglutinativ til bo‘lgani tufayli, bir so‘zning morfologik shakllari ko‘p bo‘ladi. Shuning uchun kontekstual disambiguatsiya o‘zbek tilidagi morfologik tahlilni takomillashtirish uchun asosiy shartlardan biridir.

Kontekstual disambiguatsiya jarayonini tahlil qilishdan oldin disambiguatsiya tushunchasiga ta’rif berish lozim. Disambiguatsiya (lot. *ambiguus* – “noaniq”) – bu til birliklarining noaniqligini bartaraf etish, ya’ni aniq ma’no yoki shaklni tanlash jarayoni.

Tilshunoslikda uning uch asosiy darajasi mavjud:

1. Leksik disambiguatsiya – so‘z ma’nosini aniqlash.
2. Morfologik disambiguatsiya – so‘zning grammatik shaklini aniqlash.
3. Sintaktik disambiguatsiya – gap tuzilmasidagi noaniqlikni bartaraf etish.

Tadqiqot doirasida o‘rganilayotgan morfologik disambiguatsiya – so‘zning grammatik kategoriyasini (so‘z turkumi, son, kelishik, zamon, shaxs va son kabi) aniqlash jarayonidir. Bu jarayon quyidagi bosqichlarda amalga oshiriladi:

1-jadval.

Morfologik disambiguatsiya bosqichlari

Bosqich	Mazmuni
1. Morfologik tahlil	So‘zning barcha mumkin bo‘lgan grammatik tahlillari aniqlanadi
2. Kontekstual baholash	Har bir variant atrofdagi so‘zlar asosida tahlil qilinadi
3. Eng mos variantni tanlash	Eng mos ma’no yoki shakl avtomatik ravishda belgilanadi

Misol sifatida, “yoz” so‘zini oladigan bo‘lsak, u quyidagi ma’nolarda keladi:

Yoz (fe’l) – *kitobni yoz*. Ya’ni, “Men *kitobni yozdim*”.

Yoz (ot) – *yoz keldi*. Ya’ni, “*Yoz issiq bo‘ldi*”.

Kontekstual model asosidagi disambiguatsiya esa, so‘zning ma’nosi faqat o‘zi bilan emas, balki atrofdagi so‘zlar bilan bog‘liq bo‘ladi. Shu sababli,

zamonaviy NLP modellarida disambiguatsiya kontekst asosida avtomatik o'rganiladi. Mazkur holatda, modellarning ishlash prinsipi quyidagi tartibda amalga oshiriladi:

1. Har bir so'z vektor (raqamli koordinata) shaklida ifodalanadi.
2. Model so'zning oldingi va keyingi so'zlar bilan semantik bog'liqligini o'rganadi.
3. Kontekst asosida eng to'g'ri ma'no tanlanadi.

Misol sifatida, “men ot minib bordim” (ot), “dushmanni ot” (fe'l) va boshqa shakllarda kelishi mumkin. Ikki o'rinda ham ot so'zi qo'llanilmoqda, lekin turli ma'nolarni ifodalamoqda.

O'zbek tilida kontekstual disambiguatsiya muammolari. O'zbek tilidagi NLP tizimlarida kontekstual disambiguatsiya hali to'liq avtomatlashtirilmagan va ushbu jarayonning o'ta murakkab. Hozirgi kundagi asosiy muammolar sifatida quyidagilarni keltirish mumkin:

- so'z shakllarining ko'pligi va o'xshashligi;
- affikslar kombinatsiyasining murakkabligi;
- sintaktik tartibning moslashuvchanligi (so'z tartibi erkinligi);
- raqamli korpuslarning cheklanganligi (ma'lumotlar yetishmasligi).

Ammo oxirgi yillarda bu yo'nalishda O'zbek WordNet, UzMorph, UzBERT kabi loyihalar orqali sezilarli siljishlar kuzatilmoqda.

Yuqorida keltirilgan muammolarning oldini olish maqsadida, quyidagi usullardan foydalanish maqsadga muvofiq:

Qoidalarga asoslangan (Rule-based) yondashuv. Bunda tilshunoslar tomonidan tuzilgan qoidalar asosida ma'no tanlanadi. Uning asosiy kamchiligi sifatida katta mehnat talab qilishi va kontekstga moslashuvchan emasligini keltirish mumkin.

Statistik yondashuv. Mazkur yondashuvda model quriladi va u so'zlarning birgalikda qo'llanilish chastotasiga qarab ma'noni keltirib chiqaradi.

Neyron tarmoqlarga asoslangan (deep learning) yondashuv. Yuqorida keltirilgan qoidalar, statistik yondashuvlar va bazalar asosida model yaratiladi.

Zamonaviy modellar sifatida BERT, GPT, RoBERTani keltirish mumkin. U soʻzning atrofidagi kontekstni chuqur tahlil qilib, kontekstual vektorlar yordamida disambiguatsiya qiladi. Hozirgi kunda bu yondashuv eng aniq natija bermoqda.

Kontekstual disambiguatsiya juda murakkab jarayon boʻlib, uning quyidagi muammoli sabablari mavjud:

1. Agglutinativ til xususiyati. Oʻzbek tilida soʻz ildiziga qoʻshimchalar, yasalma va hosila qoʻshimchalari, turli grammatik kategoriyalar (kelishik, son, zamon, h.k.) qoʻshilib soʻz shakllanadi. Bu qoʻshimchalar bir nechta grammatik maʼnolarni berishi mumkin. Masalan, qoʻshimchalar uzaytirilgan boʻlgan soʻzda, uni ajrata olmaydigan holatlarga duch kelinadi.

2. Yigʻma soʻzlar, frazeologizmlar, frazalar. Soʻz frazalarda yoki iboralarda alohida maʼnolaridan farqli ishlaydigan hollarda tahlil qilish qiyinlashadi (idiomatik foydalanish, figurativ maʼno, frazeologizmlar). Matematik yoki lugʻatga kiritilmagan shakllar alohida kontekst talab qiladi.

3. Ellipsis (yozma yoki ogʻzaki nutqda baʼzi soʻzlar tushirilishi). Kontekstda maʼnosi yoʻq lekin qoʻshimcha sifatida kelgan soʻzlar morfologik aniqlikni zaiflashtiradi.

4. Qoʻshimchalar va hosilalar koʻpligi va ularning kombinatsiyalari. Soʻzga bir nechta qoʻshimchalar kelishi mumkin (masalan, raqam, shaxs, zamon, qoʻshimcha), ularning joylashuvi, turli kelishma loyihalari turli shakllarga olib keladi.

5. Fonetik oʻzgarishlar va assimilyatsiya. Soʻz hosil qilish jarayonida yoki fonetik muhit taʼsiri ostida qoʻshimchalar yoki ildiz shakllari biroz oʻzgarishi mumkin, bu esa avtomatik tahlilchilarda xatolarga olib keladi.

6. Maʼlumot yetishmasligi va korpus cheklovlari. Ayrim korpuslarning yetarli darajada boʻlmasligi, turli shevalar yoki janrlarni (adabiy, ogʻzaki, ilmiy, gazetalar) qamrab olmasligi – bu disambiguatsiya imkoniyatlarini cheklaydi.

7. Polisemlik (soʻzning bir nechta maʼnosi boʻlishi). Masalan, bir soʻz turlicha maʼnoga ega boʻlishi mumkin (misol: “bosh (tana aʼzosi) yoki “bosh” (rahbar)). Kontekst boʻlmasa, togʻri tahlil qilish qiyin boʻladi.

8. Omonimlik – maʼnosi turlicha boʻlgan soʻzlarni talaffuzi va yozilishi bir xil boʻlishidir. Masalan, talaffuzi va yozilishi bir xil, ammo maʼnosi butunlay farq qiladi. “oq” – (feʼl – suv oqishi) va “oq” – (sifat – oq rang). Bunday soʻzlarning maʼnosi tahlilda kontekst asosida aniqlanadi .

9. Qoʻshimchalardagi noaniqlik (ambiguitet). Baʼzan bir xil shaklli qoʻshimcha bir nechta grammatik vazifani bajaradi (masalan, “-gan” qoʻshimchasi oʻtgan zamon yoki sifatdosh shakli yasavchi vazifasida keladi) Bunday holatlar disambiguatsiya jarayonini talab qiladi.

Ushbu muammolarni bartaraf etish uchun quyidagi yoʻllardan foydalanilishi mumkin:

1. Kontekstga asoslangan NLP modellaridan foydalanish va ularni takomillashtirish:

- BERT, XLM-R, mBERT kabi tillararo transfer va oʻzbek tilida pre-trening qilingan modellardan foydalanish;
- konkret morfologik vazifalar uchun maxsus qatlamlar (morfologik taggerlar) qoʻshish.

2. Annotatsiyalangan korpusni kengaytirish:

- turli janrlar, shevalar, ogʻzaki nutq, yozma matn, ijtimoiy media matnlari uchun xilma-xil misollar;
- morfologik annotatsiyalar, POS-tagging, lemmatizatsiya, hosila va hosilalar kombinatsiyalari, fonetik variantlar.

3. Qoidalarga asoslangan + mashinani oʻrganishga boʻlgan yondashuvlarda mujassamlashgan yondashuv: Qoidalarga asoslangan analizatorlarni mashinani oʻrganish modellari bilan birlashtirish: qoidalar kontekst asosida yuzaga keladigan ambiguitetlarni dastlabki kamaytiradi, model esa statistik xususiyatlar asosida aniqlikni oshiradi.

4. Disambiguatsiya uchun kontekstni kengaytirish:

- soʻz atrofidagi soʻzlar, kelajak yoki oldingi gaplar, sintaktik tuzilma, semantik rol, gap tuzilishi faktorlaridan keng foydalanish;

- boshqaruvchi-xizmat maʼlumotlaridan foydalanish.

5. Konstruktsiyalar va frazeologizmlarni alohida modellashtirish:

- frazeologik birliklar, idiomatik iboralarni aniqlash, va korpusda markirovka qilish;

- lugʻatlar va leksik resurslarni kengaytirish: polisemantik soʻz birikmalari, olmoshlarning barcha koʻrinishlari va hosila shakllari roʻyxatini tuzish;

- fonetik va morfofonetik qoidalarini kiritish: qoʻshimchalar bilan bogʻliq fonetik tafovutlar, assimilyatsiya yoki konsonant oʻzgarishlarini hisobga oluvchi model.

6. Baholash va metrikalarni aniq belgilash:

- morfologik analizator natijalarini baholash uchun test toʻplamlari, annotatsiyalangan misollarni ekspert tilshunoslar yordamida tuzish;

- xatolik tahlili: qayerda va qaysi tipdagi disambiguatsiya xatolari koʻp sodir boʻlayotganini aniqlash.

Kontekstual disambiguatsiya – tabiiy tilni avtomatik tahlil qilish jarayonida soʻz yoki shaklning toʻgʻri maʼnosini tanlashga qaratilgan muhim bosqichdir. Oʻzbek tili morfologiyasida bu jarayon feʼl va ot shakllarining farqlanishi, koʻplik va egalik affikslari aniqlanishi hamda soʻz tartibining semantik taʼsirini toʻgʻri tushunishga xizmat qiladi. Kelajakda oʻzbek tilidagi sunʼiy intellekt tizimlarini mukammallashtirish uchun kontekstual disambiguatsiyani neyron tarmoqlar asosida modellashtirish zarur.

FOYDALANILGAN ADABIYOTLAR:

1. Abdurakhmonova N., Alishe I., Sayfulleyeva R. MorphUz: Morphological analyzer for the Uzbek language //2022 7th International Conference on Computer Science and Engineering (UBMK). – IEEE, 2022. – pp. 61-66.

2. Axmedova X. Turli soʻz turkumlari orasidagi omonimiyani aniqlovchi matematik modellar //Science and innovation. – 2022. – T. 1. – №. B7. – p. 393-400.
3. Shahobiddinovna R.M. Soʻzlarning lugʻaviy va kontekstual maʼnosi //FARS International Journal of Education, Social Science & Humanities. – 2025. – T. 13. – №. 5. – pp. 678-681.
4. Parsing C. Speech and language processing //Power Point Slides. – 2009. – P. 20.