

BADIIY ASARLAR KORPUSIDA OMONIM SHAKLLAR TAHLILI

Hamroyeva Shahlo Mirdjonovna,

Filologiya fanlari doktori, professor v.b.

shaxlo.xamrayeva@navoiy-uni.uz

Gavharshodbegim Xudoynazarova,

Kompyuter lingvistikasi yo'nalishi talabasi

xudoynazarovagavhar@gmail.com

ToshDO'TAU

Annotatsiya. Ushbu maqolada badiiy korpus yaratish va uning Sketch Engine tizimida tahlil qilinishi muhokama qilinadi. Maqolada qidiruv tizimida qo'shimchalar bilan bog'liq muammolar va omonimik shakllarning turli holatlari aniqlanadi. Xususan, qo'shimchalar bordek ko'rinib, aslida qo'shimchaga ega bo'lmagan so'zlar, qo'shimcha qotib qolgan so'zlar, so'z oxirida muammolarga olib keladigan holatlar va kontekstual o'zgarishlar kabi istisno holatlar tahlil qilinadi. Maqola morfoanalizatorning qidiruv tizimidan foydalanishda uchraydigan muammolarni hal qilish uchun yangi metodologiyalarni ishlab chiqishga yordam beradi, hamda badiiy matnlardagi omonim qo'shimchalar bilan bog'liq kompleks holatlarni chuqurroq tushunishga imkon beradi.

Annotation. This article discusses the creation of a literary corpus and its analysis in the Sketch Engine system. It identifies issues related to affixes in the search system and various cases of homonymous forms. In particular, the article examines exceptional cases such as words that appear to have affixes but actually do not, words with fossilized affixes, problematic word endings that cause errors, and contextual changes. The study contributes to the development of new methodologies to address challenges encountered when using a morphological analyzer in the search system and provides a deeper understanding of complex cases involving homonymous affixes in literary texts.

Аннотация. В данной статье рассматривается создание художественного корпуса и его анализ в системе Sketch Engine. В статье выявляются проблемы, связанные с аффиксами в поисковой системе, а также различные случаи омонимичных форм. В частности, анализируются такие исключительные случаи, как слова, которые внешне выглядят как имеющие аффиксы, но фактически их не содержащие, слова с застывшими аффиксами, проблемные окончания слов, приводящие к ошибкам, а также контекстуальные изменения. Статья способствует разработке новых методологий для решения проблем, возникающих при использовании морфоанализатора в поисковой системе, а также позволяет глубже понять сложные случаи омонимичных аффиксов в художественных текстах.

Kalit soʻzlar. *Badiiy korpus, metodologiya, kompleks holat, omonim qoʻshimchalar.*

Badiiy korpus yaratish va ularni tahlil qilish – tilshunoslikda zamonaviy texnologiyalardan foydalanishning muhim qismidir. Korpuslarni tahlil qilishda, lingvistik jihatdan ayniqsa qiyin boʻlgan omonimlar va qoʻshimchalar oʻzining alohida oʻrniga ega. Bu maqolada biz Sketch Engine dasturi orqali yaratgan badiiy korpusda uchragan omonimik shakllar va qoʻshimchalar bilan bogʻliq tahlilga toʻxtalamiz. Oʻtkir Hoshimovning “Ikki karra ikki besh”, “Dunyoning ishlari”, “Hurkitilgan tush”, Tohir Malikning “Shaytanat”, “Jinoyatning uzun yoʻli”, va Omina Shelinkoʻgʻlining “Iskanja” asarlarini raqamlashtirib, dasturga yuklash orqali biz turli xil omonim soʻzlarni va ularning qoʻshimchalarini tahlil qilishni amalga oshirdik.

Badiiy asarlar korpusini ihlab chiqish va tahlil qilish

Sketch Engine[8] – bu tilshunoslik tahlili uchun ishlatiladigan onlayn dastur boʻlib, foydalanuvchilarga matnlarini avtomatik ravishda tahlil qilish imkonini beradi. Sketch Engine – lingvistik korpuslarni avtomatik tahlil qilish uchun ishlab chiqilgan ilgʻor platformalardan biri boʻlib, turli tillardagi katta hajmdagi matnlarni qayta ishlash imkonini beradi. Ushbu tizim quyidagi afzalliklarga ega:

- 1) matnlardan avtomatik ravishda soʻz shakllari va ularning kontekstini ajratib olish;
- 2) omonimik birliklarning real foydalanish chastotasini tahlil qilish;
- 3) morfologik va sintaktik xususiyatlarni aniqlash orqali til tadqiqotlarini osonlashtirish.

Ayniqsa, Sketch Engine tizimi tilshunoslar va sunʼiy intellekt dasturchilari uchun katta ahamiyatga ega boʻlib, u lingvistik model va algoritmlarni yaratishda qoʻllaniladi.

Tahlil qilinayotgan badiiy korpusda 6 ta asar mavjud:

1. Oʻ.Hoshimov. Ikki karra ikki besh.
2. Oʻ.Hoshimov. Dunyoning ishlari.
3. Oʻ.Hoshimov. Hurkitilgan tush.
4. Tohir Malik. Shaytanat.
5. Tohir Malik. Jinoyatning uzun yoʻli.
6. Omina Shelinkoʻgʻli. Iskanja (tarjima asar).

Ushbu asarlar adabiyotda keng qo'llaniladigan leksik va semantik elementlarga boy bo'lib, ularning har biri badiiy uslubni o'rganishda muhim manbadir.

Bugungi kunda lingvistika va sun'iy intellekt tahlillarida badiiy korpus yaratish muhim vosita sifatida ahamiyat kasb etadi. Til birliklarini chuqur o'rganish va ularning kontekstga bog'liq holda qanday o'zgarishini aniqlashda korpus lingvistikasi samarali usul bo'lib, tahlil jarayonini aniq va tizimli olib borishga imkon beradi. Badiiy korpuslar nafaqat lingvistik tahlil, balki stilistika, semantika va pragmatika bo'yicha tadqiqotlarda ham muhim ahamiyat kasb etadi. Sun'iy intellekt yordamida korpus tahlilini amalga oshirish, matnlarni avtomatik kategoriyalarga ajratish, omonimik shakllarning kontekstga bog'liq holda qanday ishlatilishini aniqlash kabi imkoniyatlarni beradi. Bu esa tilshunoslik tadqiqotlarining yanada kengayishiga xizmat qiladi.

Badiiy asarlar korpusining lingvistik tadqiqotlardagi o'rni

Badiiy asarlar korpusini yaratish bugungi kunda lingvistika va sun'iy intellekt tahlillari uchun muhim vosita ekanligini alohida ta'kidlash lozim. Korpus lingvistikasining asosiy maqsadi til birliklarining real kontekstdagi qo'llanilishini o'rganishdir. Badiiy matnlar esa tilning eng tabiiy va stilistik jihatdan boy shakllaridan biri bo'lib, leksik-semantik tadqiqotlar uchun juda muhim manba hisoblanadi. Korpus lingvistikasida badiiy matnlarning ahamiyati quyidagilarda namoyon bo'ladi:

Til birliklarining kontekstga bog'liq holda qanday ishlatilishini tahlil qilish. Til birliklarining kontekstga bog'liq holda qanday ishlatilishini tahlil qilish – lingvistikaning asosiy yo'nalishlaridan biri bo'lib, so'z va iboralar ma'nosining kontekstga qarab o'zgarishini o'rganish bilan bog'liq. Bunday tahlil ayniqsa, semantika, pragmatika, stilistika, va tabiiy tilni qayta ishlash (NLP) sohalarida muhim o'rin tutadi.

Stilistik va pragmatik tadqiqotlar uchun real ma'lumotlar bazasini yaratish. Stilistik va pragmatik tadqiqotlar uchun real ma'lumotlar bazasi yaratish – zamonaviy lingvistika va sun'iy intellekt sohalarida muhim yo'nalishlardan biri hisoblanadi. Bunday bazalar tabiiy tilning murakkab jihatlarini o'rganish, tahlil qilish va avtomatlashtirish uchun asos bo'lib xizmat qiladi.

Sun'iy intellektga asoslangan tabiiy tilni qayta ishlash (NLP) tizimlari uchun modellarni shakllantirish. Tabiiy tilni qayta ishlash (Natural Language Processing – NLP) sun'iy intellektning eng ilg'or yo'nalishlaridan biri bo'lib, u inson tilini tushunish, tahlil qilish va yaratish imkonini beradi. NLP modellarini shakllantirish murakkab jarayon bo'lib, u ma'lumotlarni yig'ish, tozalash, o'qitish va optimallashtirish kabi bosqichlarni o'z ichiga oladi.

1. Omonim so'zlar va qo'shimchalar tahlili

Tilshunoslikda omonimiya masalasi eng muhim va murakkab yo'nalishlardan biri hisoblanadi. Chunki bir xil shakldagi so'zlar turli kontekstlarda turlicha ma'noga ega bo'lishi mumkin. Sketch Engine tizimi orqali badiiy korpusda qo'shimchalar va omonimik shakllarni qidiruv asosida tahlil qilindi. Qidiruv tizimida ba'zi omonim so'zlar va qo'shimchalar bilan bog'liq muammolar aniqlandi. Bu muammolarni quyidagi 3 ta guruhga ajratish mumkin:

Qo'shimcha bordek ko'rinadi, lekin qo'shimchasi bo'lmagan so'zlar. Ba'zi so'zlar tashqi ko'rinishda qo'shimchaga o'xshasa ham, aslida qo'shimcha bilan bog'lanmagan. Masalan: Chinni (chin+ni) – bu so'zda **-ni** shakli mavjud, lekin morfem jihatdan qo'shimcha emas. Pechka (pech+ka) – bu so'zda ham tashqi ko'rinishda qo'shimcha mavjud kabi ko'rinadi, lekin bu so'zda **-ka** bo'g'in, qo'shimcha emas. Bunga o'xshash bir qancha so'zlar mavjud: (1-2-3-4-jadvallar)

1-jadval. Tushum kelishigi (-ni) shakli omonimiyasi

№	So'z	-ni
1.	Chinni	chin+ni
2.	Ayni	ay+ni
3.	Jinni	jin+ni
4.	Mini	
5.	Luqmoni	
6.	Poni	

2-jadval. Jo'nalish kelishigi (-ka) shakli omonimiyasi

№	So'z	-ka
1.	Pechka	pech+ka
2.	Yekla	yel+ka
3.	Sirka	sir+ka
4.	Erka	er+ka
5.	Tikka	tik+ka
6.	To'nka	to'n+ka

3-jadval. Jo 'nalish kelishigi (-qa) shakli omonimiyasi

№	So'z	-qa
1.	Halqa	hal+qa
2.	Orqa	or+qa
3.	Qisqa	qis+qa
4.	Buqa	bu+qa
5.	Boshqa	bosh+qa
6.	Aloqa	alo+qa
7.	Yoqa	yo+qa
8.	Loyqa	loy+qa
9.	Antiqqa	anti+qa
10.	Manqa	man+qa

4-jadval. Chiqish kelishigi (-dan) shakli omonimiyasi

№	So'z	-dan
1.	Aftidan	aft+i+dan
2.	Jumladan	jumla+dan
3.	Boshqatdan	boshqa+t+dan
4.	Zimdan	
5.	Obdan	
6.	Qo'qisidan	

Qo'shimcha qotib qolgan (leksemalashgan) so'zlar. Ba'zi so'zlarda qo'shimchalar o'zining asl ma'nosini yo'qotgan, so'z tarkibida qotib qolgan. Misol uchun: Boshqacha (boshqa+cha) – bu so'z semantik va leksik jihatdan o'zgarib, tarkibiy qismi qotib qolgan so'zlardan biridir. Bu holatda “-cha” qo'shimchasi o'zining asl ma'nosini yo'qotgan va yangi ma'noga ega bo'lgan. Asosan ravish so'z turkumidagi so'zlarda qo'shimchalar qotib qoladi va ular ajratilmaydi. Bunga o'xshash bir qancha so'zlar mavjud: (Qarang: 5-jadval)

5-jadval. Tarkibida qo'shimcha qotib qolgan so'zlar

№	So'z	qo'shimcha
1.	Boshqacha	boshqa+cha
2.	Birdan	bir+dan
3.	Ataydan	atay+dan
4.	Nihoyatdan	nihoyat+da
5.	Kunda	kun+da
6.	Birga	bir+ga
7.	Chalqanchasiga	chalqancha+si+ga

8.	Birdaniga	bir+dan+i+ga
9.	Yonlamasiga	yonlama+si+ga
10.	Qatorasiga	qator+a+si+ga
11.	Bekorga	bekor+ga

So‘z qidiruvdagi noaniqlikni keltirib chiqaruvchi shakllar. O‘zbek tilida shunday so‘zlar borki, ularning oxirgi bo‘g‘ini qo‘shimchalar bilan omonimlik hosil qiladi va bu qidiruv tizimida muammoga olib keladi. Masalan: Jonona (jon+ona) – bu so‘zning oxirgi bo‘g‘ini “ona” qo‘shimchasi bilan omonimlik hosil qila oladi va qidiruv tizimida tarkibida “ona” bo‘lgan so‘zlarni ham qo‘shimcha sifatida chiqariladi. Bunga o‘xshash bir qancha so‘zlar mavjud (Qarang: 6-jadval).

6-jadval. So‘z qidiruvda muammoga uchraydigan so‘zlar.

№	So‘z	Ona
1.	Mastona	mast +ona
2.	Kamtarona	kamtar +ona
3.	Ayyorona	ayyor +ona
4.	Ojizona	ojiz +ona
5.	Mardona	mard +ona
6.	Ahmoqona	Ahmoq +ona

2. Tizimdagi muammo va yechimlar

Omonim so‘z va qo‘shimchalarning muammoli holatlari Sketch Engine tizimida ba’zi natijalarning noto‘g‘ri ko‘rsatishiga olib keldi. Bu, ayniqsa, qo‘shimchalar o‘zining leksik shaklini o‘zgartirgan yoki qotib qolgan so‘zlarda uchradi. Agar bu muammolarni hal qilmasak, qidiruv tizimi noto‘g‘ri natijalar ko‘rsatadi. Muammolarni bartaraf etish uchun quyidagi yondashuvlar talab etiladi:

1. So‘zlarni tahlil qilishda semantik qatlamlarni hisobga olish: qo‘shimchalar va omonim so‘zlarning ma’nolarini tushunish va ularning kontekstga qarab qanday ishlashini aniqlash.

2. Qo‘shimcha va so‘z shakllarini tekshirish: korpusdagi har bir so‘zni alohida tahlil qilib, qaysi qo‘shimchalar ishlatilishini aniqlash va ularga asoslangan alohida qidiruv mexanizmlarini ishlab chiqish.

3. Istisno holatlar va muammolar

Istisno holatlar mavjud bo‘lib, ular quyidagi uch guruhda tahlil qilindi:

1. Qo‘shimcha bordek ko‘rinadi lekin qo‘shimchasi bo‘lmagan so‘zlar.
2. Qo‘shimcha qotib qolgan so‘zlar.
3. So‘zni qidirganda oxirida muammoga uchraydigan so‘zlar.

Xulosa. Sketch Engine tizimidan foydalangan holda badiiy korpusda omonimik shakl va qo'shimchalarni tahlil qilish tilshunoslik tadqiqotlarida muhim o'rin tutadi. Biroq ba'zi muammolarni, ayniqsa, qo'shimchalar va omonimlar bilan bog'liq holatlarni aniqlash va tuzatish zarur. Yangi tahlil metodologiyalarini ishlab chiqish, so'zning morfem tarkibini to'g'ri aniqlash va qidiruv tizimlarini optimallashtirish badiiy matnlar va lingvistik tahlilni yanada samarali qilishga yordam beradi. Ushbu tadqiqot natijalari, ayniqsa, avtomatlashtirilgan tarjima tizimlari, matnlarni avtomatik qayta ishlash, nutqni aniqlash va sun'iy intellekt asosidagi lingvistik dasturlarni rivojlantirishda amaliy ahamiyat kasb etadi. Omonim shakllarning kontekstual o'zgarishlarini chuqur o'rganish orqali tilga oid dasturlarni yanada mukammallashtirish imkoniyati yuzaga keladi.

Foydalanilgan adabiyotlar:

1. Hoshimov O. Dunyoning ishlari: Qissa va hikoyalar. – Toshkent: Yangi asr avlodi, 2015.
2. Malik T. Shaytanat. 1-qism. – Islom.uz ma'rifat markazidan olindi. Manba: www.ziyouz.com, 2007.
3. Malik T. Jinoyatning uzun yo'li. – PDF versiyasi. Manba: www.ziyouz.com.
4. Hoshimov O. Hurkitilgan tush: Hikoyalar. – Toshkent: Ziyonashr, nashr yili ko'rsatilmagan.
5. Shenliko'g'li O. Iskanja. – Toshkent, 2021. (Nashr tili: o'zbek, kirill).
6. Hoshimov O. Ikki karra ikki – besh. – Toshkent: Nurli dunyo nashriyot uyi, 2024.
7. Abjalova M. Korpus lingvistikasi: Uslubiy qo'llanma. – Toshkent: Bookmany Print, 2022.
8. Sketch Engine. Sketch Engine – corpus analysis platform. Manba: <https://www.sketchengine.eu>.
9. Yusupova R. D. Sketch Engine platformasida korpuslar yaratishning amaliy tadqiqi // *E-Conference Series*. – ISSN: 2835-5326. – 46-bet. Manba: econferenceseries.com.
10. Zaxarov V., Mengliyev B., Xamroyeva Sh. Korpus lingvistikasi: Korpus tuzish va undan foydalanish. – Globe Edit