

e-ISSN 3060-5059

DOI: 10.67227



IJTIMOYIY-GUMANITAR SOHADA ILMIY-INNOVATSION TADQIQOTLAR

ILMIY METODIK JURNAL

Scientific and Innovative Research in
Social and Humanitarian Studies

Vol. 3 • Issue 9

2026



Founded in 2024



publishscience.uz



publish_science_journal@mail.ru



PEER REVIEWED • OPEN ACCESS

O‘ZBEK TILIDAGI OMONIM LKSEMALARNI KONTEKST ASOSIDA AVTOMATIK FARQLASHNING LINGVISTIK VA METODOLOGIK ASOSLARI

Xolmurodova Iroda Axtamovna

O‘zbekiston Milliy universiteti, magistrant.

e-mail: ixolmurodova703@gmail.com ORCID: <https://orcid.org/0009-0006-6021-3894>

Annotatsiya.

Maqolada o‘zbek tilidagi omonim leksemalarni kontekst asosida avtomatik farqlash muammosi lingvistik va metodologik jihatdan tahlil qilingan. Tadqiqot o‘zbek tilidagi ayrim omonim leksemalar misolida lotin va kirill yozuvlarida ikkita kontekstual model — UzBERT va XLM-R yordamida olib borilgan. Har bir leksemaning turli ma’nolariga oid kontekstlar kosinus o‘xshashligi asosida taqqoslanib, ichki va tashqi o‘xshashlik ko‘rsatkichlari hamda ular orasidagi ajratish farqi hisoblangan. Natijalar xalqaro miqyosda tan olingan F1 baholash ko‘rsatkichi orqali qo‘shimcha ravishda tasdiqlangan: UzBERT oltita sinov holatining beshtasida 80% dan yuqori F1 ko‘rsatkichiga erishgan, XLM-R esa 20% dan 100% gacha keskin beqarorlikni namoyon etgan. Ushbu natijalar tilga xos oldindan o‘qitilgan modellarning ko‘p tilli modellarga nisbatan o‘zbek tilidagi omonimiyani farqlashda barqarorroq ishlashini ko‘rsatadi.

Kalit so‘zlar: omonimiya, omonim leksema, WSD, kontekstual tahlil, semantik differensiallash, UzBERT, XLM-R, F1.

СИНТАКСИЧЕСКАЯ НЕЙТРАЛИЗАЦИЯ: СЕМАНТИКО-ФУНКЦИОНАЛЬНЫЙ ПОДХОД

Холмуродова Ирода Ахтамовна

Национальный университет Узбекистана, магистрант.

Аннотация.

В статье с лингвистической и методологической точек зрения анализируется проблема автоматического различения омонимичных лексем узбекского языка на основе контекста. Исследование проведено на примере отдельных омонимичных лексем узбекского языка в латинской и кириллической графике с использованием двух контекстуальных моделей — UzBERT и XLM-R. Контексты, соответствующие различным значениям каждой лексемы, сопоставлялись на основе косинусного сходства; были рассчитаны показатели внутреннего и внешнего сходства, а также разница между ними. Полученные результаты дополнительно подтверждены с помощью международно признанной метрики F1: UzBERT достиг показателя F1 выше 80% в пяти из шести тестовых случаев, тогда как XLM-R продемонстрировал значительную нестабильность с результатами от 20% до 100%. Полученные данные свидетельствуют о том, что языково-специфичные предварительно обученные модели обеспечивают более стабильные результаты при различении омонимии в узбекском языке по сравнению с многоязычными моделями.

Ключевые слова: омонимия, омонимичная лексема, WSD, контекстуальный анализ, семантическая дифференциация, UzBERT, XLM-R, F1.

SYNTACTIC NEUTRALIZATION: A SEMANTIC-FUNCTIONAL PERSPECTIVES

Xolmurodova Iroda Axtamovna

National University of Uzbekistan, Master’s Student.

Abstract.

This article analyzes the problem of automatic context-based disambiguation of homonymous lexemes in the Uzbek language from linguistic and methodological perspectives. The study was conducted on selected Uzbek homonymous lexemes represented in both Latin and Cyrillic scripts using two contextual models, UzBERT and XLM-R. Contexts corresponding to different meanings of each lexeme were compared using cosine similarity, and intra-class and inter-class similarity scores, as well as the separation gap between them, were calculated. The results were further validated using the internationally recognized F1 evaluation metric: UzBERT achieved an F1 score above 80% in five out of six test cases, whereas XLM-R demonstrated substantial instability, with scores ranging from 20% to 100%. These findings indicate that language-specific pretrained models provide more stable performance in distinguishing homonymy in Uzbek than multilingual models.

Keywords: homonymy, homonymous lexeme, WSD, contextual analysis, semantic

differentiation, UzBERT, XLM-R, F1.

Kirish. Sun'iy intellekt texnologiyalarining jadal rivojlanishi tabiiy tilni avtomatik qayta ishlash (NLP) tizimlarini zamonaviy raqamli infratuzilmaning ajralmas qismiga aylantirdi. Biroq bunday tizimlarning sifati, avvalo, ularning tilning eng nozik semantik hodisalarini qay darajada to'g'ri tushunishiga bog'liq. Shunday hodisalardan biri omonimiya, ya'ni shakli bir xil, biroq semantik jihatdan o'zaro bog'liq bo'lmagan birliklarning mavjudligi bo'lib, u so'z ma'nosini avtomatik aniqlash — Word Sense Disambiguation (WSD) sohasining eng murakkab va hozirgacha to'liq yechilmagan muammolaridan biri sifatida xorijiy tadqiqotlarda ham e'tirof etilgan. Kontekstda qo'llangan so'zning to'g'ri ma'nosini oldindan belgilangan ma'nolar inventaridan tanlash (WSD) mashina tarjimasini, axborot qidiruvi va sentiment tahlili kabi NLP vazifalarining sifatiga bevosita ta'sir qiladi [18; 18–21-b.].

Ushbu mavzu xalqaro miqyosda ham nihoyatda faol rivojlanib borayotgan tadqiqot yo'nalishlaridan biri hisoblanadi. Omonimiya munosabatiga asoslangan semantik klasterlash g'oyasi va uni yirik kontekstual modellar (BERT, RoBERTa, DeBERTa, XLM-R) yordamida tekshirish 2024-yilda inglizcha WordNet resursi asosida amalga oshirilgan [19; 924–938-b.]. Omonimlik munosabatlarini hikoya kontekstida baholashga bag'ishlangan SemEval-2026 Task 5 kabi xalqaro tanlovlar esa aynan joriy yilda o'tkazilmoqda. Demak, mazkur muammo hatto resurslarga boy tillar uchun ham eskirgan yoki yakunlangan masala emas, balki hozirgi kunda shakllanib borayotgan va faol ilmiy muhokama qilinayotgan tadqiqot yo'nalishidir. Xorijiy tadqiqotlarda ushbu masala turli tillar misolida faol o'rganilmoqda: Xitoyda X. Zhou va boshqalar (2024), J. Du, F. Qi va M. Sun (2019), Turkiyada B. İlgen (2015), G. Ercan va O. Yıldız (2018), Rossiyada esa T. Kaushinis va boshqalar (2015) hamda O. Xomichivich va boshqalar (2013) tomonidan tegishli tadqiqotlar olib borilgan.

O'zbek tilida bu muammo agglyutinativ morfologiya, so'z turkumlari orasidagi grammatik yaqinlik va omonim shakllarning yuqori chastotasi bilan kuchayadi. Shu sababli tokenning yuzaki shakli yetarli emas, so'z turkumi, morfologik tarkib, sintaktik qurshov va kengroq kontekst birgalikda hisobga olinishi talab etiladi. O'zbek kompyuter lingvistikasida omonimlikni aniqlash masalasi qoidaviy-morfologik tasniflash, statistik-ehtimoliy klassifikatsiya darajasida bir qancha tadqiqotchilar tomonidan o'rganilgan. Kengroq yo'nalishda N. Abduraxmonova (2024) o'zbek tili elektron korpusi va tezaurus standartlari bo'yicha, B. Mansurov va A. Mansurov (2021) esa UzBERT kontekstual modeli bo'yicha tadqiqotlar olib borgan.

Hozirda o'zbek kompyuter lingvistikasida mavjud ishlarning umumiy cheklovi shundaki, ular morfologik yoki statistik klassifikatsiya darajasida to'xtaydi va bir xil yoki turli xil so'z turkumiga tegishli, semantik jihatdan bog'lanmagan ma'nolarni ajratish uchun alohida qatlam taklif qilmaydi. Xalqaro tajriba ko'rsatgan kontekstual klasterlash bosqichi esa o'zbek tiliga hali qo'llanilmagan. Mazkur tadqiqotda esa o'zbek tilidagi omonim leksemalarni kontekst asosida avtomatik farqlashning lingvistik va metodologik asoslari ishlab chiqildi va tajribadan o'tkazildi.

FOYDALANILGAN ADABIYOTLAR SHARHI

So'nggi yillarda aynan Word Sense Disambiguation (so'z ma'nosini aniqlash, farqlash) masalasi xalqaro miqyosda o'rganilayotgan dolzarb mavzulardan biridir. Xorijiy olimlardan R. Naviglining maqolasida (2009) WSD sohasining kompleks, tizimli sharhi berilib, asosiy tushunchalar shakllantirilgan. A. Vaswani va boshqalar (2017) transformer arxitekturasini taklif qilgan va u zamonaviy barcha kontekstual modellarning asosi hisoblanadi, J. Devlin va boshq. (2019) BERT modelini yaratib, kontekstual til modellashtirishda burilish yasagan tadqiqotchi olimlar. A. Raganato va boshq. (2017) WSD tizimlarini baholash uchun yagona, standart freymvork taklif qilgan, A. Conneau va boshq. (2020) tadqiqotlarida 100 dan ortiq tilni qamrab oluvchi ko'p tilli kontekstual model yaratilgan, L. Proietti va boshq. (2024) WordNet ma'nolarini omonim klasterlarga birlashtirib, BERT/RoBERTa/ELECTRA/DeBERTa modellarining ularni farqlash qobiliyatini o'rgangan.

J. Du, F. Qi va M. Sun (2019) BERTni bevosita so'z ma'nosini aniqlash vazifasiga qo'llagan. X. Zhou va boshq. (2024) xitoycha ma'no bildiruvchi belgilarning tarkibidan foydalanib BERTni semantik jihatdan boyitgan. G. Ercan va O. Yıldız (2018) turk tili uchun semantik model baholash resursini (so'z o'xshashligi, bog'liqligi) yaratgan. B. İlgen (2015) turk tili uchun nazoratli va nazoratsiz WSD usullarini doktorlik dissertatsiyasida qiyoslagan. T. Kaushinis va boshq. (2015) lug'aviy ko'p ma'nolilikni hal qilishning turli matematik-algoritmik usullarini (neyron tarmoq, AdaBoost, Bayes, klasterlash) tizimli sharhlagan. O. Xomchivich, S. Ribin, I. Anichkin (2013) rus nutq sintezi tizimida matnni normallashtirish va omonimiyani hal qilishda lingvistik tahlildan foydalangan. O'zbek kompyuter lingvistikasida ham ushbu tadqiqot uchun fundamental asoslar mavjud. N. Abduraxmonova (2018) mashina tarjimasini tizimlari uchun lingvistik ta'minot masalalarini o'rgangan hamda o'zbek tilining elektron korpusini yaratish uchun kompyuter modellarini ishlab chiqqan.

TADQIQOT METODOLOGIYASI

Ushbu tadqiqotda o'zbek tilidagi omonim leksemalarni sun'iy intellektga asoslangan kontekstual modellar (UzBERT, XLM-R) doirasida semantik jihatdan farqlash asosiy maqsad qilib olingan. Shu bois tadqiqotda tahlil va sintez, qiyosiy-tipologik hamda hisoblash (eksperimental) metodlari kompleks ravishda qo'llanildi. Har bir metod uch qatlamli farqlash sxemasini shakllantirishga xizmat qiladi, ya'ni omonim nomzodni morfologik aniqlash, semantik differensiallash va kontekstual ma'noni tanlash. Ushbu metodologik yondashuv natijasida o'zbek tiliga moslashtirilgan, kontekst asosida ishlaydigan omonim farqlash modelini ishlab chiqish masalalari ko'rib chiqildi.

TAHLIL VA NATIJALAR

Ma'nolarni farqlash (Word Sense Disambiguation, WSD) tabiiy tilni qayta ishlashning (NLP) asosiy vazifalaridan biri bo'lib, kontekstdagi so'zga to'g'ri ma'no (sema)ni biriktirishni maqsad qiladi. Biroq an'anaviy WSD tizimlari WordNetni asosiy ma'nolar to'plami sifatida qo'llaydi va ko'pincha so'z ma'nolarining semantik farqlarini o'ta chuqur farqlaydi, bu esa zamonaviy NLPda WSD tizimlarining ortiqcha murakkablashishiga va amaliy qo'llanilishining pasayishiga olib kelishi mumkin. Haqiqatan ham, hozirgi oldindan o'qitilgan til modellari (PLM) ma'nolarni farqlashga qodirdek ko'rinadi, biroq bu qobiliyat qay darajada ishonchli ekanligi aniq emas [19; 924-938-b.]. So'z ma'nosini aniqlashda eng muhim muammolardan biri bu omonimiyani bartaraf etish, ya'ni yozilishi yoki talaffuzi bir xil, lekin ma'nosi turlicha bo'lgan so'zlarni to'g'ri farqlash masalasidir. Biz aynan shu muammoni bartaraf etish maqsadida omonimlarning ma'nosini kontekst asosida avtomatik aniqlash modelini taklif etamiz. Taklif etilayotgan model uch bosqichdan iborat: birinchidan, kontekstga joylashtirilgan omonim leksema oldindan o'qitilgan kontekstual model (UzBERT yoki XLM-R) orqali vektorli tasvirga (embedding) aylantiriladi; ikkinchidan, ushbu vektor har bir ma'no guruhi uchun oldindan hisoblangan markaziy vektor (centroid) bilan kosinusli o'xshashlik asosida taqqoslanadi; uchinchidan, eng yuqori o'xshashlikka ega markazga tegishli ma'no guruhi modelning yakuniy bashorati sifatida tanlanadi (nearest-centroid klassifikatsiya).

Hozirgacha xalqaro miqyosda o'tkazilgan shunga o'xshash tadqiqotlar PLMlarning bunday farqlashni sezilarli aniqlik bilan amalga oshira olishini ko'rsatgan bo'lsa-da, bu natijalar asosan ingliz tili va WordNet kabi keng ko'lamli resurslarga tayangan holda olingan. Bu natijalarning o'zbek tili kabi struktur jihatdan farqli va morfologik jihatdan boy tilga qanchalik amal qilishi esa hozircha ochiq savol bo'lib qolmoqda [15; 387-443-b.]. Ushbu ilmiy-metodologik bo'shliqni bartaraf etish maqsadida biz ushbu tadqiqotda o'zbek tili uchun maxsus oldindan o'qitilgan UzBERT modeli hamda ko'p tilli XLM-R modelining omonim leksemalarni kontekst asosida farqlash borasidagi qiyosiy salohiyatini dastlabki sinovdan o'tkazdik.

Ushbu dastlabki sinov uch xil tabiatga ega omonim leksema asosida o'tkazildi: "ot" (uch ma'noli: hayvon, ism hamda otmoq fe'li), "yuz" (turli xil

Model	Yozuv	Hayvon ichida	Ism ichida	Otmoq ichida	Hayvon↔Ism	Hayvon↔Otmoq	Ism↔Otmoq	Ajratish farqi
UzBERT	lotin	0.8150	0.7354	0.7841	0.7367	0.7077	0.7178	-0.0014
UzBERT	kirill	0.7407	0.6058	0.6051	0.6135	0.5072	0.4489	-0.0084
XLM-R	lotin	0.9916	0.9926	0.9827	0.9896	0.9817	0.9861	-0.0070
XLM-R	kirill	0.9893	0.9825	0.9874	0.9858	0.9866	0.9850	-0.0041

1-jadval. "Ot" leksemasi bo'yicha to'liq natijalar

1- jadvalda "ot" leksemasining hayvon, ism va otmoq ma'nolariga oid gaplar asosida hisoblangan ichki va tashqi (↔) o'xshashlik qiymatlari hamda ularning ajratish farqi keltirilgan. Navabatdagi 2 va 3-jadvallarda ham aynan shu farqlar yuz va qo'y omonim leksemalari misolida ikki model o'rtasida qiyoslangan holda hisoblab chiqilgan:

Model	Yozuv	Yuz (bet) ichida	Yuz (100) ichida	Bet↔100	Ajratish farqi
UzBERT	lotin	0.7788	0.7410	0.6954	0.0456
UzBERT	kirill	0.6369	0.5789	0.4915	0.0873
XLM-R	lotin	0.9871	0.9873	0.9856	0.0015
XLM-R	kirill	0.9923	0.9922	0.9920	0.0002

2-jadval. "Yuz" leksemasi bo'yicha to'liq natijalar

Model	Yozuv	Hayvon ichida	Fe'l ichida	Hayvon↔Fe'l	Ajratish farqi
UzBERT	lotin	0.8808	0.8612	0.8438	0.0174
UzBERT	kirill	0.7273	0.7714	0.6837	0.0436
XLM-R	lotin	0.9963	0.9966	0.9952	0.0011
XLM-R	kirill	0.9901	0.9924	0.9919	-0.0017

3-jadval. "Qo'y" leksemasi bo'yicha to'liq natijalar

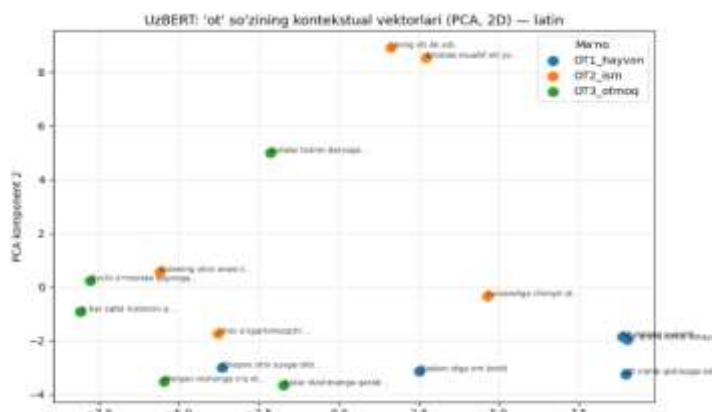
Model	Leksema	Yozuv	Ajratish farqi
UzBERT	ot	lotin	-0.0014
UzBERT	ot	kirill	-0.0084
UzBERT	yuz	lotin	0.0456
UzBERT	yuz	kirill	0.0873
UzBERT	qo'y	lotin	0.0174
UzBERT	qo'y	kirill	0.0436
XLM-R	ot	lotin	-0.0070
XLM-R	ot	kirill	-0.0041
XLM-R	yuz	lotin	0.0015
XLM-R	yuz	kirill	0.0002
XLM-R	qo'y	lotin	0.0011
XLM-R	qo'y	kirill	-0.0017

4-jadval. 1,2,3-jadvallarning umumiy xulosasi (Ajrata olish farqi)

Umumiy olti sinov holatining to'rttasida (yuz-lotincha, yuz-kirillcha, qo'y-lotincha, qo'y-kirillcha) UzBERT ijobiy ajratish farqiga erishdi (+0,0174 dan +0,0873 gacha), XLM-R esa barcha holatlarda deyarli nolga yaqin yoki manfiy natija berdi. XLM-R 100 dan ortiq tilni bitta umumiy parametrlar to'plamida qamrab olishga majbur bo'lgani uchun har bir alohida til uchun nozik semantik farqlarni saqlab qola olmaydi, UzBERT esa faqat o'zbek tiliga ixtisoslashgani sababli turli ma'nolarni ancha yaxshi ajratadi. Ushbu natija so'nggi xalqaro tadqiqotlar bilan ham mos keladi [7; 25–26-b.], ya'ni so'z ma'nosini aniqlash vazifasida kichik, maqsadli o'qitilgan modellar ko'lami kattaroq, umumiy til modellaridan ustun bo'lishi mumkin. Bu fikr o'zbek tili misolida UzBERT va XLM-R o'rtasidagi taqqoslashda ham o'z tasdig'ini topdi.

Alohida e'tiborga loyiq yana bir jarayon shuki, yozuv tizimining ta'siri "yuz" va "qo'y" leksemalarida UzBERT modelida kirillcha yozuvdagi matnlar lotincha yozuvdagi matnlarga nisbatan izchil ravishda yuqoriroq ajratish farqini berdi. Bu B. Mansurov va A. Mansurov (2021) o'z ishlarida qayd etgan faktga to'g'ridan-to'g'ri mos keladi. Mualliflarning o'zlari ta'kidlaganidek, UzBERT faqat kirillcha matnlar asosida o'qitilgan. "Ot" leksemasida esa bu jarayon kuzatilmadi (ikkala model uchun ham manfiy natija), bu ehtimol "hayvon" va "ism" ma'nolarining grammatik shakl jihatidan o'ta o'xshashligi bilan bog'liq istisno holat bo'lishi mumkin. Ammo kuzatishlar natijasida shu aniq bo'ldiki, modellarning ajrata olish qobiliyati bir xil so'z turkumi doirasida emas, balki ikki ma'no orasidagi semantik masofaga bog'liq: qanchalik uzoq mavzular (tibbiyot va qishloq xo'jaligi "qovoq" so'zi misolida) bo'lsa, shunchalik oson farqlanadi, qanchalik yaqin va bir-biriga aralashib ketadigan mavzular (jonli mavjudot, nomlash) bo'lsa, shunchalik qiyin farqlanadi.

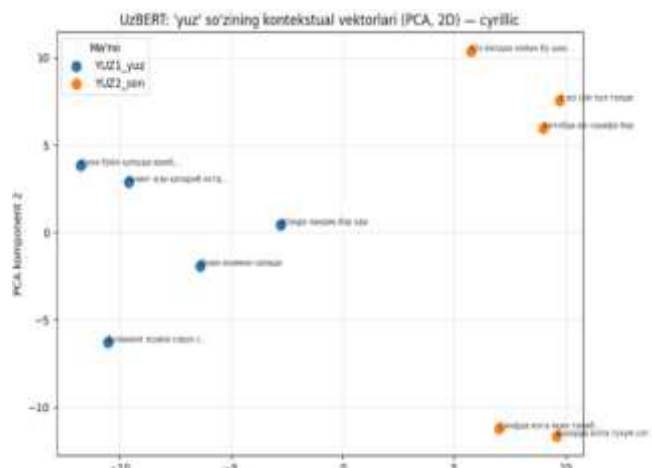
Sonli natijalarni vizual tarzda tasdiqlash maqsadida E. Reif va boshq. (2019) tomonidan taklif etilgan yondashuvga asoslanib, tanlangan uch holat uchun so'zlarning kontekstga bog'liq, ko'p o'lchovli vektorli tasvirlari (embeddinglari) bosh komponentlar tahlili (PCA) yordamida ko'rish uchun qulay bo'lgan ikki o'lchovli tekislikka tushirildi (proyeksiyalandi) [21; 8592–85600-b.].



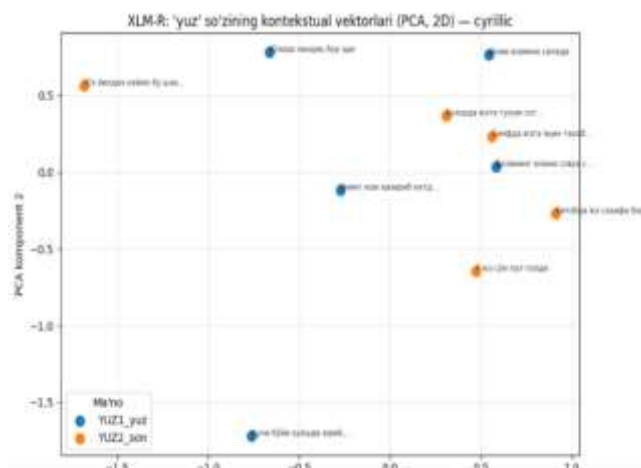
1-rasm. UzBERT: "ot" so'zining kontekstual vektorlari (PCA, 2D) – lotin

1-rasmda ko'rsatilganidek, "hayvon" (ko'k rangda) ma'nosiga oid nuqtalar boshqa ikkisidan qisman ajralib tursa-da, "ism" (to'q sariq rangda) va "otmoq" (yashil rangda) ma'nolariga oid nuqtalar bir-biriga kuchli aralashib, aniq chegaralanmagan holda joylashgan bu uch ma'noli omonimlarni farqlashning ikki ma'noli holatlarga nisbatan qanchalik murakkabroq ekanligini vizual tarzda ko'rsatadi va ajratish farqi ko'rsatkichini (–0,0014) vizual jihatdan tasdiqlaydi.

Buning aksicha, 2-rasmda "yuz" leksimasining insonning yuzi va son ma'nolariga tegishli nuqtalar ikki aniq, bir-biridan chetlashgan to'daga guruhlangan, bu esa UzBERT modelining ushbu ikki ma'noni kontekst orqali izchil farqlay olishini vizual tarzda ko'rsatadi:



2-rasm. UzBERT: “yuz” so’zining kontekstual vektorlari (PCA, 2D) — kirill



3-rasm. XLM-R: “yuz” so’zining kontekstual vektorlari (PCA, 2D) — kirill

Aynan shu leksema 3-rasmda “yuz” kirillcha XLM-R modeli orqali tasvirlangan bo’lib, nuqtalar orasida deyarli hech qanday vizual ajratish kuzatilmaydi. Bu ikkala model o’rtasidagi farqni bir xil leksema va bir xil yozuv misolida to’g’ridan-to’g’ri qiyoslash imkonini beradi.

Natijalarni yanada mustahkamlash maqsadida quyida F1 baholash ko’rsatkichi ham hisoblab chiqildi.

F1 baholash ko’rsatkichi quyidagicha aniqlanadi:

$$F1 \text{ (aniqlik)} = 2 \cdot (\text{Aniqlik} \cdot \text{Eslash}) / (\text{Aniqlik} + \text{Eslash})$$

bu yerda:

$$\text{Aniqlik} = TP / (TP + FP), \text{Eslash} = TP / (TP + FN)$$

F1 baholash ko’rsatkichi qanday vazifani bajaradi? Bu ko’rsatkich modelning har bir gapni to’g’ri ma’no guruhiga qanchalik ishonchli ajrata olganini yagona songa jamlab beradi, ya’ni model tomonidan berilgan barcha bashoratlarning necha foizi haqiqiy ma’noga mos kelganini ko’rsatadi. Boshqacha aytganda, u modelning umumiy klassifikatsiya sifatini qanchalik kam xato bilan (kam yolg’on musbat, kam yolg’on manfiy) ishlaganini bitta, tushunarli va solishtirsa bo’ladigan ko’rsatkichga aylantiradi [21; 8592–8600-b.].

Ushbu tadqiqot sharoitida:

- TP (to’g’ri musbat) — modelning muayyan ma’no guruhiga to’g’ri tegishli deb bashorat qilgan holatlari soni;
- FP (yolg’on musbat) — boshqa ma’noga tegishli bo’lgan, biroq model tomonidan noto’g’ri shu guruhga kiritilgan holatlar soni;
- FN (yolg’on manfiy) — aslida shu ma’no guruhiga tegishli bo’lgan, biroq model boshqa guruhga bashorat qilgan holatlar soni.

F1 baholash ko’rsatkichi har bir sinov holati uchun chetlab qoldirish (leave-one-out) asosida [14; 1137–1143-b.] va eng yaqin markazga (nearest-centroid) klassifikatsiya yondashuvi yordamida hisoblab chiqildi [16; 506-b.]. Natijalar quyidagi jadvalda keltirilgan:

5-jadval. Omonim leksemalar bo’yicha F1 baholash ko’rsatkichi natijalari

Model	Leksema	Yozuv	F1 (aniqlik)
UzBERT	ot	lotin	67%
UzBERT	ot	kirill	93%
UzBERT	yuz	lotin	90%
UzBERT	yuz	kirill	100%
UzBERT	qo’y	lotin	80%
UzBERT	qo’y	kirill	100%
XLM-R	ot	lotin	73%
XLM-R	ot	kirill	40%
XLM-R	yuz	lotin	80%
XLM-R	yuz	kirill	40%
XLM-R	qo’y	lotin	100%
XLM-R	qo’y	kirill	20%

Natijalar shuni ko’rsatadiki, UzBERT modeli olti sinov holatining beshtasida 80% dan yuqori F1 baholash ko’rsatkichiga erishdi, XLM-R esa ancha beqaror bo’lib, 20% dan 100% gacha keskin tebranib turdi. Bu o’zbek tiliga maxsus o’qitilgan modelning ko’p tilli modelga nisbatan ushbu vazifada barqaror ishlashini tasdiqlaydi.

XULOSA

Xulosa o'rnida shuni aytish mumkinki, o'zbek tilida omonimiyani aniqlash masalasi murakkab jarayon. Omonimiyani aniqlashda ularni turli guruhlariga ajratish va ularni farqlovchi omillarga asosan turli usullardan foydalanish maqsadga muvofiq. O'tkazilgan tadqiqotda o'zbek tilidagi omonim leksemalarni kontekst asosida farqlash uchun uch qatlamli (morfologik, semantik, kontekstual) model taklif etilib, uning dastlabki shartini tekshirish maqsadida "ot", "yuz" va "qo'y" leksemalari asosida UzBERT va XLM-R modellari ishtirokida hisoblash tajribasi o'tkazildi. Olti sinov holatining to'rttasida "yuz" va "qo'y" leksemalarining ikkala yozuvida o'zbek tiliga maxsus o'qitilgan UzBERT modeli ma'nolarni izchil farqlagani holda, ko'p tilli XLM-R buni deyarli uddalay olmadi, bu esa tilga xos modellarning agglyutinatib til uchun ustunligini tasdiqladi. Ushbu xulosa F1 baholash ko'rsatkichi orqali ham mustahkam tasdiqlandi: UzBERT olti sinov holatining beshtasida 80% dan yuqori F1ga erishdi, XLM-R esa 20% dan 100% gacha keskin beqarorlik ko'rsatdi. Shuningdek, kirillcha yozuvdagi matnlarning lotincha yozuvdagi matnlarga nisbatan izchil yuqori natija berishi UzBERT modelining faqat kirillcha yozuvdagi korpusda o'qitilganligi bilan bog'liq ekanligi aniqlandi. Natijalar PCA orqali vizual tarzda ham tasdiqlanib, o'zbek kompyuter lingvistikasida hozirgacha yetishmagan bo'g'inni empirik dalillar bilan to'ldirdi.

FOYDALANILGAN ADABIYOTLAR

1. Abduraxmonova N. *Mashina tarjiması tizimlari uchun lingvistik ta'minot masalalari*. – Toshkent: Fan, 2018. – 184 b.
2. Abduraxmonova N. O'zbek tili tezaurusini yaratishda standart me'yorlar // *Til va adabiyot ta'limi*. – Toshkent, 2024. – № 3. – B. 45–49.
3. Abduraxmonova N., Adali E., Shamieva N. Semantic Relationships Between Synset Adjectives for English-Uzbek Bilingual Thesaurus // *2025 10th International Conference on Computer Science and Engineering (UBMK)*. – 2025. – P. 305–310. – DOI: [10.1109/UBMK67458.2025.11206834](https://doi.org/10.1109/UBMK67458.2025.11206834). Maqola va mualliflar tarkibi tasdiqlandi. [ResearchGate](https://www.researchgate.net/publication/391811111) URL: [DOI orqali maqola](https://doi.org/10.1109/UBMK67458.2025.11206834)
4. Abduraxmonova N., Rajabova K. Stages of Creating Uzbek Language Multimedia Corpus // *2025 10th International Conference on Computer Science and Engineering (UBMK)*. – 2025. – P. 1628–1633. – DOI: [10.1109/UBMK67458.2025.11206927](https://doi.org/10.1109/UBMK67458.2025.11206927). Sahifalar va konferensiya ma'lumotlari tasdiqlandi. [DOI](https://doi.org/10.1109/UBMK67458.2025.11206927) URL: [IEEE/DOI sahifasi](https://doi.org/10.1109/UBMK67458.2025.11206927)
5. Abduraxmonova N., Shamsiyeva G. Context-Based Multilingual Translation Technology: On the Example of the Paratranslator Platform // *2025 10th International Conference on Computer Science and Engineering (UBMK)*. – 2025. – P. 1756–1760. – DOI: [10.1109/UBMK67458.2025.11206971](https://doi.org/10.1109/UBMK67458.2025.11206971). [ResearchGate](https://www.researchgate.net/publication/391811111) URL: [DOI orqali maqola](https://doi.org/10.1109/UBMK67458.2025.11206971)
6. Agostini A., Usmanov T., Khamdamov U., Abdurakhmonova N., Mamasaidov M. UZWORDNET: A Lexical-Semantic Database for the Uzbek Language // *Proceedings of the 11th Global Wordnet Conference*. – 2021. – P. 8–19. – DOI: [10.18653/v1/2021.gwc-1.2](https://doi.org/10.18653/v1/2021.gwc-1.2). [ACL Anthology](https://www.aclweb.org/anthology/2021.gwc-1.2) URL: [ACL Anthology](https://www.aclweb.org/anthology/2021.gwc-1.2)
7. Basile P., Siciliani L., Musacchio E., Semeraro G. Exploring the Word Sense Disambiguation Capabilities of Large Language Models // *arXiv preprint arXiv:2503.08662*. – 2025. Maqola 2025-yil 11-martda taqdim etilgan. [arXiv](https://arxiv.org/abs/2503.08662) URL: [arXiv](https://arxiv.org/abs/2503.08662)
8. Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzmán F., Grave E., Ott M., Zettlemoyer L., Stoyanov V. Unsupervised Cross-Lingual Representation Learning at Scale // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. – 2020. – P. 8440–8451. – DOI: [10.18653/v1/2020.acl-main.747](https://doi.org/10.18653/v1/2020.acl-main.747). [ACL Anthology](https://www.aclweb.org/anthology/2020.acl-main.747) URL: [ACL Anthology](https://www.aclweb.org/anthology/2020.acl-main.747)
9. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // *Proceedings of NAACL-HLT 2019*. – Minneapolis, Minnesota, 2019. – P. 4171–4186. – DOI: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423). [ACL Anthology](https://www.aclweb.org/anthology/N19-1423) URL: [ACL Anthology](https://www.aclweb.org/anthology/N19-1423)
10. Du J., Qi F., Sun M. Using BERT for Word Sense Disambiguation // *arXiv preprint arXiv:1909.08358*. – 2019. [arXiv](https://arxiv.org/abs/1909.08358) URL: [arXiv](https://arxiv.org/abs/1909.08358)
11. Ercan G., Yıldız O. T. AnlamVer: Semantic Model Evaluation Dataset for Turkish – Word Similarity and Relatedness // *Proceedings of the 27th International Conference on Computational Linguistics*. – Santa Fe, New Mexico, USA, 2018. – P. 3819–3836. [ACL Anthology](https://www.aclweb.org/anthology/I18-1036) URL: [ACL Anthology](https://www.aclweb.org/anthology/I18-1036)
12. Hadiwinoto C., Ng H. T., Gan W. C. Improved Word Sense Disambiguation Using Pre-Trained Contextualized Word Representations // *Proceedings of EMNLP-IJCNLP 2019*. – Hong Kong, China, 2019. – P. 5297–5306. – DOI: [10.18653/v1/D19-1533](https://doi.org/10.18653/v1/D19-1533). [ACL Anthology](https://www.aclweb.org/anthology/D19-1533) URL: [ACL Anthology](https://www.aclweb.org/anthology/D19-1533)
13. İlgen B. *Türkçe sözcük anlam belirsizliği giderme: Doktora tezi*. – İstanbul: İstanbul Teknik Üniversitesi, 2015. – 165 s.
14. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection // *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI 1995)*. – 1995. – Vol. 2. – P. 1137–1145. [ML Anthology](https://www.aclweb.org/anthology/IJCAI1995-1137) URL: [IJCAI rasmiy PDF](https://www.aclweb.org/anthology/IJCAI1995-1137)
15. Loureiro D., Rezaee K., Pilehvar M. T., Camacho-Collados J. Analysis and Evaluation of Language Models for Word Sense Disambiguation // *Computational Linguistics*. – 2021. – Vol. 47, No. 2. – P. 387–443. – DOI: [10.1162/coli_a_00405](https://doi.org/10.1162/coli_a_00405). [DOI](https://www.mitpress.mit.edu/publications/doi/10.1162/coli_a_00405) URL: [MIT Press / DOI](https://www.mitpress.mit.edu/publications/doi/10.1162/coli_a_00405)
16. Manning C. D., Raghavan P., Schütze H. *Introduction to Information Retrieval*. – Cambridge: Cambridge University Press, 2008. – 482 p. [Stanford NLP](https://nlp.stanford.edu/) URL: [Stanford kitob sayti](https://nlp.stanford.edu/)
17. Mansurov B., Mansurov A. UzBERT: Pretraining a BERT Model for Uzbek // *arXiv preprint arXiv:2108.09814*. – 2021. [arXiv](https://arxiv.org/abs/2108.09814) URL: [arXiv](https://arxiv.org/abs/2108.09814)

18. Navigli R. Word Sense Disambiguation: A Survey // *ACM Computing Surveys*. – 2009. – Vol. 41, No. 2. – Article 10. – P. 1–69. – DOI: [10.1145/1459352.1459355](https://doi.org/10.1145/1459352.1459355). [ResearchGate](https://www.researchgate.net/publication/11171172) URL: [DOI / ACM](https://doi.org/10.1145/1459352.1459355)
19. Proietti L., Perrella S., Tedeschi S., Vulpis G., Lavallo L., Sanchietti A., Ferrari A., Navigli R. Analyzing Homonymy Disambiguation Capabilities of Pretrained Language Models // *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. – Torino, Italy, 2024. – P. 924–938. [ACL Anthology](https://aclanthology.org/2024.lrec-main.164/) URL: [ACL Anthology](https://aclanthology.org/2024.lrec-main.164/)
20. Raganato A., Camacho-Collados J., Navigli R. Word Sense Disambiguation: A Unified Evaluation Framework and Empirical Comparison // *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. – Valencia, Spain, 2017. – P. 99–110. [ACL Anthology](https://aclanthology.org/E17-1.10/) URL: [ACL Anthology](https://aclanthology.org/E17-1.10/)
21. Reif E., Yuan A., Wattenberg M., Viégas F. B., Coenen A., Pearce A., Kim B. Visualizing and Measuring the Geometry of BERT // *Advances in Neural Information Processing Systems*. – 2019. – Vol. 32. – P. 8592–8600. [proceedings.neurips.cc](https://proceedings.neurips.cc/paper_files/paper/2019/file/Reif-Visualizing-and-Measuring-the-Geometry-of-BERT.pdf) URL: [NeurIPS rasmiy sahifasi](https://neurips.cc/paper_files/paper/2019/file/Reif-Visualizing-and-Measuring-the-Geometry-of-BERT.pdf)
22. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention Is All You Need // *Advances in Neural Information Processing Systems*. – 2017. – Vol. 30. – P. 5998–6008. [proceedings.neurips.cc](https://proceedings.neurips.cc/paper_files/paper/2017/file/Vaswani-Attention-Is-All-You-Need.pdf) URL: [NeurIPS rasmiy sahifasi](https://neurips.cc/paper_files/paper/2017/file/Vaswani-Attention-Is-All-You-Need.pdf)
23. Zhou X., Huang H., Chi Z., Ren M., Gao Y. RS-BERT: Pre-training Radical Enhanced Sense Embedding for Chinese Word Sense Disambiguation // *Information Processing & Management*. – 2024. – Vol. 61, No. 4. – Article 103740. – DOI: [10.1016/j.ipm.2024.103740](https://doi.org/10.1016/j.ipm.2024.103740).
- Muhim tuzatish:** sizdagi 103712 noto'g'ri; Elsevier rasmiy yozuvida maqola raqami 103740. [ScienceDirect](https://www.sciencedirect.com/science/article/pii/S0306457324000374) URL: [ScienceDirect / DOI](https://doi.org/10.1016/j.ipm.2024.103740)
24. Каушинис Т. В., Кириллов А. Н., Коржицкий Н. И., Крижановский А. А., Пилинович А. В. и др. Обзор методов и алгоритмов разрешения лексической многозначности: введение // *Труды Карельского научного центра РАН*. – Петрозаводск, 2015. – № 8. – С. 3–23.
25. Хомицевич О. Г., Рыбин С. В., Аничкин И. М. Использование лингвистического анализа для нормализации текста и снятия омонимии в системе синтеза русской речи // *Известия высших учебных заведений. Приборостроение*. – 2013. – Т. 56, № 2. – С. 45–50. [Открытые книги](https://www.itmo.ru/openbooks/monographs/2013/2013_02_045-50.pdf) URL: [ITMO OpenBooks maqola sahifasi](https://www.itmo.ru/openbooks/monographs/2013/2013_02_045-50.pdf)