



KO‘P TILLI SUN‘IY INTELLEKT TIZIMLARIDA SEMANTIK NOMUVOFIQLIK MUAMMOLARI

Abdumalikova Mushtariybegim Mashxurbek qizi

O‘zbekiston davlat jahon tillari universiteti

Xorijiy til va adabiyot fakulteti 1-bosqich talabasi.

Annotatsiya. Mazkur maqolada ko‘p tilli sun‘iy intellekt tizimlarida uchraydigan semantik nomuvofiqlik muammolari lingvistik va texnologik nuqtai nazardan tahlil qilinadi. Polisemiya, omonimiya, frazeologik birliklar, madaniy kontekst va pragmatik ma‘no tafovutlari sun‘iy intellekt modellarining aniqligiga qanday ta‘sir ko‘rsatishi ilmiy asosda yoritiladi. Shuningdek, kichik resursli tillar, xususan o‘zbek tili misolida real tarjima va generativ model xatolari ko‘rib chiqiladi. Tadqiqot davomida semantik buzilishlarni kamaytirishning strategik yo‘nalishlari taklif etiladi. Natijalar ko‘p tilli sun‘iy intellekt tizimlarini rivojlantirishda chuqur lingvistik integratsiya zarurligini ko‘rsatadi.

Kalit so‘zlar: ko‘p tilli sun‘iy intellekt, semantik nomuvofiqlik, polisemiya, frazeologiya, pragmatika, neyron tarjima, kichik resursli tillar.

Annotation. This article provides a theoretical and practical analysis of semantic inconsistency in multilingual artificial intelligence systems. It examines how polysemy, homonymy, idiomatic expressions, cultural context, and pragmatic meaning differences affect the performance of AI models. Particular attention is given to low-resource languages, especially Uzbek, through real examples of translation and generative model errors. The study proposes strategic approaches to reduce semantic distortions and emphasizes the necessity of deeper linguistic integration in multilingual AI development.



Keywords: multilingual artificial intelligence, semantic inconsistency, polysemy, pragmatics, neural translation, low-resource languages.

Аннотация. В статье рассматриваются теоретические и практические аспекты семантической несогласованности в многоязычных системах искусственного интеллекта. Анализируются явления полисемии, омонимии, фразеологии, культурного контекста и прагматических различий. Особое внимание уделяется языкам с ограниченными ресурсами, в частности узбекскому языку. Предлагаются стратегические направления минимизации семантических искажений. Результаты исследования подчеркивают необходимость углубленной лингвистической интеграции при разработке многоязычных ИИ-систем.

Ключевые слова: многоязычный искусственный интеллект, семантическая несогласованность, прагматика, нейронный перевод, языки с ограниченными ресурсами.

Zamonaviy axborot texnologiyalari davrida matnli ma'lumotlarning hajmi misli ko'rilmagan darajada oshdi. Ijtimoiy tarmoqlar, elektron hujjatlar, yangiliklar portallari va boshqa platformalar har kuni millionlab matnli xabarlarni yaratmoqda. Statistik ma'lumotlarga ko'ra, 2025-yilga kelib global ma'lumotlar hajmining 175 zettabaytga yetishi prognoz qilinmoqda, bunda matnli ma'lumotlar muhim qismni tashkil etadi. Ushbu ma'lumotlarni qo'lda tahlil qilish imkonsiz bo'lib, avtomatlashtirilgan tahlil usullariga ehtiyoj ortib bormoqda. Matnlarni semantik analiz qilish – bu matnning ma'nosini aniqlash, uning ichki mantiqiy tuzilmasini tushunish va kontekstga asoslangan xulosalar chiqarish jarayoni bo'lib, ushbu muammoni hal qilishda asosiy vosita sifatida qaralmoqda. Matnlarni semantik tahlil qilishning dolzarbligi bir qator omillarga bog'liq. Birinchidan, ijtimoiy tarmoqlarda tarqalayotgan axborotni tezkor tahlil qilish orqali jamoatchilik fikrini monitoring



qilish, soxta xabarlarining tarqalishini oldini olish va marketing strategiyalarini optimallashtirish imkoniyati paydo bo'ladi. Ikkinchidan, o'zbek tili kabi kam resursli tillar uchun semantik analiz vositalarini ishlab chiqish milliy axborot makonini rivojlantirishda muhim qadamdir. Uchinchidan, sun'iy intellektning rivojlanishi bilan matn tahlili avtomatik tarjimada, suhbat agentlarida va qidiruv tizimlarida keng qo'llanilmoqda. An'anaviy matn tahlil usullari, masalan, kalit so'zlarga asoslangan statistik yondashuvlar yoki qoida asosidagi algoritmlar, murakkab kontekstlarni tushunishda cheklovlarga ega. Masalan, "Men bu kitobni yaxshi ko'raman" va "Men bu kitobni yaxshi ko'rmayman" jumlaridagi "yaxshi" so'zi bir xil bo'lsa-da, umumiy ma'no butunlay boshqacha. Shu sababli, kontekstni chuqur tahlil qila oladigan zamonaviy usullarga ehtiyoj sezilmoqda. O'zbek tilida esa bu muammo yanada dolzarb, chunki mavjud modellar asosan ingliz tili kabi resurslarga boy tillar uchun optimallashtirilgan. Ushbu maqolaning asosiy maqsadi matnlarni semantik analiz qilishning zamonaviy usullarini, xususan mashinaviy o'qitish va tabiiy tilni qayta ishlash (NLP) texnologiyalarini o'rganish, ularning o'zbek tilidagi matnlarga qo'llanilishini sinovdan o'tkazish va samaradorligini baholashdan iborat.

Sun'iy intellekt texnologiyalarining rivojlanishi tabiiy tilni qayta ishlash sohasida mutlaqo yangi imkoniyatlarni ochib berdi. Ayniqsa transformator arxitekturasi asosidagi ko'p tilli modellar turli tillarda matn yaratish, tarjima qilish va semantik tahlil qilish jarayonini sezilarli darajada tezlashtirdi. Biroq inson tilining murakkab semantik tabiati sun'iy tizimlar uchun hanuzgacha to'liq hal etilmagan muammo bo'lib qolmoqda. Til faqat so'zlar yig'indisi emas, balki madaniyat, tarix, ijtimoiy tajriba va kognitiv modelning ifodasidir. Shuning uchun ko'p tilli sun'iy intellekt tizimlarida ma'no ifodalanishi ko'pincha yuzaki statistik moslik darajasida qolib ketadi.



Semantik nomuvofiqlik deganda bir til birligining boshqa tilga o‘tkazilishi yoki generatsiya qilinishi jarayonida uning asl ma’nosining buzilishi, torayishi yoki noto‘g‘ri talqin qilinishi tushuniladi. Bu jarayon ayniqsa ko‘p tilli modellar uchun murakkabdir, chunki ular turli til tizimlaridagi semantik maydonlarni yagona matematik fazoda birlashtirishga harakat qiladi. Natijada ma’no qatlamlari o‘rtasida mos kelmasliklar yuzaga keladi.

Tilshunoslik nuqtai nazaridan ma’no bir necha qatlamdan iborat bo‘lib, denotativ, konnotativ va pragmatik komponentlarni o‘z ichiga oladi. Sun’iy intellekt modellarining asosiy kuchi denotativ ma’noni aniqlashda namoyon bo‘ladi. Ammo konnotativ va pragmatik ma’no ko‘pincha kontekstga, madaniy tajribaga va nutq vaziyatiga bog‘liq bo‘lganligi sababli algoritmik tahlil uchun murakkablik tug‘diradi. Masalan, o‘zbek tilidagi “ko‘ngli og‘ridi” iborasi literal tarjimada fiziologik jarayonni anglatadi, biroq aslida emotsional holatni ifodalaydi. Sun’iy model kontekst yetarli bo‘lmaganda ushbu farqni ajrata olmaydi.

Ko‘p tilli sun’iy intellekt tizimlarining ishlash mexanizmi statistik ehtimollik va vektorli reprezentatsiyaga asoslanadi. Har bir so‘z yoki ibora ko‘p o‘lchovli fazoda joylashgan vektor sifatida ifodalanadi. Turli tillardagi semantik yaqin birliklar ushbu fazoda bir-biriga yaqin joylashtiriladi. Biroq har bir tilning semantik struktura modeli bir xil emas. Masalan, ingliz tilida mavjud bo‘lgan ayrim tushunchalar o‘zbek tilida kengroq yoki torroq ma’no doirasiga ega bo‘lishi mumkin. Shu sababli matematik yaqinlik har doim ham semantik ekvivalentlikni anglatmaydi.

Polisemiya ko‘p tilli modellar uchun alohida muammo hisoblanadi. Bir so‘zning bir nechta ma’noga ega bo‘lishi kontekstga asoslangan aniqlikni talab qiladi. Ingliz tilidagi “light” so‘zi yorug‘lik, yengil yoki yoqmoq ma’nosini anglatishi mumkin. Agar model kontekstual signalni yetarli darajada qayta ishlamasa, noto‘g‘ri ma’no tanlanadi. Xuddi shunday, o‘zbek tilidagi “yurdi” so‘zi



harakatni ham, amal qilishni ham anglatishi mumkin. Kontekstual chuqurlik yetishmaganda semantik xato yuz beradi.

Frazeologik birliklar va metaforalar esa semantik nomuvofiqlikning eng murakkab ko‘rinishidir. Inson tafakkuri ko‘pincha metaforik konseptlarga asoslanadi. “Hayot yo‘lga o‘xshaydi”, “vaqt puldir” kabi iboralar konseptual metafora nazariyasiga asoslanadi. Sun‘iy intellekt tizimlari esa bunday birliklarni ko‘pincha kompozitsion tahlil qiladi, ya‘ni iborani alohida so‘zlarga ajratib, literal ma‘no asosida qayta ishlaydi. Natijada “ishi yurishib ketdi” iborasi “his work started walking” shaklida tarjima qilinishi mumkin, bu esa semantik nomuvofiqlikning yaqqol misolidir.

Madaniy kontekst ham muhim omil hisoblanadi. Til madaniyatdan ajralgan holda mavjud emas. “Mahalla” tushunchasi o‘zbek ijtimoiy hayotida muayyan institutni anglatadi, biroq ingliz tilidagi “neighborhood” tushunchasi bunday ijtimoiy funksiyani to‘liq ifodalamaydi. Ko‘p tilli modellar global ma‘lumotlar asosida o‘qitilgani sababli lokal madaniy semantikani yetarlicha aks ettira olmaydi. Bu ayniqsa kichik resursli tillar uchun dolzarb muammo hisoblanadi.

Kichik resursli tillar muammosi ko‘p tilli sun‘iy intellekt tizimlarida semantik nomuvofiqlikni kuchaytiruvchi omil sifatida namoyon bo‘ladi. O‘zbek tili kabi tillar uchun annotatsiyalangan korpuslar hajmi cheklangan. Natijada model yetarlicha kontekstual naqshlarni o‘rganmaydi va umumlashtirish jarayonida xatolarga yo‘l qo‘yadi. Bu esa pragmatik va konnotativ ma‘nolarning yo‘qolishiga olib keladi.

Amaliy kuzatuvlar shuni ko‘rsatadiki, generativ modellar o‘zbek tilidagi metaforik va frazeologik birliklarni ko‘pincha literal talqin qiladi. “Ko‘zi to‘q inson” iborasi “a person with a full eye” tarzida tarjima qilinishi mumkin, holbuki uning asl ma‘nosi qanoatli, ochko‘z emas degan tushunchani bildiradi. Bunday xatolar sun‘iy intellektning konseptual semantikani chuqur anglamasligini ko‘rsatadi.



Semantik nomuvofiqlikni kamaytirish uchun lingvistik annotatsiyani kengaytirish, madaniy ontologiyalarni ishlab chiqish va kontekstual chuqur o‘qitish mexanizmlarini takomillashtirish zarur. Shuningdek, kichik resursli tillar uchun maxsus o‘qitilgan modellar yaratish ham muhim strategiya hisoblanadi. Kelajakda sun’iy intellekt tizimlari kognitiv fan, pragmatika va madaniyatshunoslik bilan integratsiyada rivojlanishi zarur bo‘ladi.

Xulosa qilib aytganda, ko‘p tilli sun’iy intellekt tizimlarida semantik nomuvofiqlik muammosi faqat texnologik emas, balki lingvistik va madaniy masaladir. Tilning ko‘p qatlamli tabiati sun’iy modellar uchun murakkab vazifa bo‘lib qolmoqda. Shu sababli kelajakdagi tadqiqotlar semantik chuqurlikni oshirish, pragmatik va madaniy komponentlarni modelga integratsiya qilishga qaratilishi lozim.

Foydalanilgan adabiyotlar:

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

2. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.

3. Shukhrat Kamalov, Diyora Absalamova, Go‘zal Absalamova, Jamila Kamalova, Farangiz Tengelova, & Munisahon Makhamedova. (2025). SEMANTIC SEARCH THROUGH VECTOR STORES: SIGNIFICANCE IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS OF NLP MODELS. Web of



Technology: Multidimensional Research Journal, 3(3), 31–39. Retrieved from <https://webofjournals.com/index.php/4/article/view/3653>

4. Schwenk, H., & Li, X. (2018). A Corpus for Multilingual Document Classification in Eight Languages. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan.

5. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.