

**SINONIMLARNI O‘QITISHDA KOMPYUTER LINGVISTIKASI VA NLP
INTEGRATSIYASI**

**Muhammad al-Xorazmiy nomidagi Toshkent axborot
texnologiyalari universiteti chet tillar kafedrası dotsenti,
f.f.f.d Saydaliyeva Gavharxon Avazovna
gavharwlu@mail.ru ID ORCID 0000-0002-1078-7818**

Annotatsiya: Ushbu maqolada NLP sistemalarida sinonimlarni o‘qitish mexanizmi lingvodidaktik va hisoblash yondashuvlari kesishmasida tahlil qilinadi. Maqsad sinonimiyani avtomatik aniqlash, guruhlash va ta’limiy vazifalarda qo‘llash uchun bosqichma-bosqich model taklif etishdir. Metodologiya korpus asosida taqsimot semantikasi, kontekstual embeddinglar va nazoratli o‘qitishni uyg‘unlashtiradi. Ilmiy yangilik sinonimlikni baholash mezonlarini didaktik darajalar bilan bog‘laydigan o‘qitish konturini asoslashdan iborat.

Kalit so‘zlar: *sinonimiya, korpus lingvistikasi, taqsimot semantikasi, kontekstual embedding, nazoratli o‘qitish, leksik resurs, lingvodidaktika*

**ИНТЕГРАЦИЯ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ И NLP В ОБУЧЕНИИ
СИНОНИМАМ**

**Ташкентского университета информационных
технологий имени Мухаммеда аль-Хорезми
Доцент кафедры иностранных языков
Сайдалиева Гавхархон Авазовна
gavharwlu@mail.ru ID ORCID 0000-0002-1078-7818**

Аннотация: В статье анализируется механизм обучения синонимам в NLP-системах на стыке лингводидактики и вычислительной лингвистики. Цель состоит в разработке поэтапной модели для автоматического выявления, группировки и использования синонимии в учебных задачах. Методология объединяет корпусную дистрибутивную семантику, контекстные эмбединги и контролируемое обучение. Научная новизна заключается в обосновании контура обучения, связывающего критерии синонимичности с дидактическими уровнями.

Ключевые слова: *синонимия, корпусная лингвистика, дистрибутивная семантика, контекстные эмбединги, контролируемое обучение, лексический ресурс, лингводидактика*

**INTEGRATION OF COMPUTATIONAL LINGUISTICS AND NLP IN TEACHING
SYNONYMS**

**Tashkent University of Information Technologies
named after Muhammad al-Khwarizmi
Associate Professor of the Foreign Languages Department
Saydaliyeva Gavharxon Avazovna
gavharwlu@mail.ru ID ORCID 0000-0002-1078-7818**

Annotation: This article examines mechanisms for training synonym knowledge in NLP systems at the intersection of language pedagogy and computational linguistics. The aim is to propose a staged model to detect, cluster, and apply synonymy in educational tasks. The methodology integrates corpus-based distributional semantics, contextual embeddings, and supervised learning. Scientific novelty lies in linking synonymy evaluation criteria to didactic proficiency levels through an explicit training loop.

Keywords: *synonymy, corpus linguistics, distributional semantics, contextual embeddings, supervised learning, lexical resources, language pedagogy*

Sinonimiya til tizimida leksik birliklar ma'nosining qisman yaqinligi, uslubiy bo'yoq, qo'llanish doirasi va nutqiy vazifa orqali farqlanishi bilan tavsiflanadi; shu sababli sinonimlarni "bir ma'noli almashtirish" sifatida soddalashtirish NLP va ta'lim amaliyotida tizimli xatolarga olib keladi. Pedagogik yo'nalishi nuqtayi nazaridan muammo faqat lingvistik tasnif masalasi emas, balki o'rganuvchi ongida sinonimik qatorni qanday shakllantirish, qaysi birlikni qaysi kommunikativ vaziyatda tanlashni o'rgatish, hamda baholashni avtomatlashtirish masalasidir. Shuning uchun NLP sistemasi orqali sinonimlarni o'qitish mexanizmi ikki qatlamli vazifa sifatida ko'riladi: birinchi qatlamda model sinonimlik munosabatini ishonchli aniqlashi va o'lchashi kerak, ikkinchi qatlamda esa olingan sinonimik bilim didaktik maqsadga mos, darajalangan, izohlanadigan va nazorat qilinadigan ta'limiy konturga joylashtirilishi zarur. Leksik ma'no va qo'llanishning kontekstga bog'liqligi haqidagi qarashlar, shuningdek, semantik yaqinlikni korpus asosida o'lchash tamoyillari ushbu tezisning nazariy poydevorini belgilaydi.

Sinonimlarni o'qitishda NLP yondashuvining asosiy ustunligi shundaki, u real nutqdan olingan korpus ma'lumotlari asosida "qaysi so'z qaysi kontekstda uchraydi" degan empirik dalilni markazga qo'yadi. Biroq korpus statistikasi o'zi yetarli emas: ikki birlik bir xil mavzu doirasida ko'p uchrashi ularning to'liq sinonimligini kafolatlamaydi, aksincha, ularning mavzudoshligi yoki kollokatsion yaqinligini ko'rsatishi mumkin. Shu bois mexanizmning birinchi bosqichi sifatida ma'lumotlarni tayyorlashda korpus muvozanati, janrlar kesimi, og'zaki va yozma matnlar nisbatini didaktik vazifaga mos tanlash talab qilinadi. O'zbek tili uchun bu jarayonda morfologik o'zgaruvchanlik, affikslar ko'pligi, yozuv variantlari hamda so'z turkumlari o'rtasidagi o'tish holatlari alohida e'tiborni talab qiladi; aks holda model sinonimlikni emas, shakliy o'xshashlikni o'rganib qoladi. Shuning uchun normalizatsiya, lemmatizatsiya va morfologik teglash orqali bir birlikning turli shakllari bitta leksik yadroga keltiriladi, keyin esa kontekst oynalari va sintaktik bog'lanishlar ajratib olinadi; bu amallar semantik o'xshashlikni "yuzaki qo'shnichilik"dan ajratishga yordam beradi.

Ikkinchi bosqich taqsimot semantikasiga tayangan holda sinonim nomzodlarni topishdan iborat. Klassik vektor modellarida so'zning ma'nosi uning qo'shni birliklari bilan ifodalanadi, natijada kosinus o'xshashligi kabi o'lchovlar orqali yaqinlik topiladi. Biroq zamonaviy NLPda kontekstual embeddinglar so'zning ma'nosini har bir kontekstda qayta hisoblaydi va ko'p ma'nolilikni qisman yechadi; bu ayniqsa o'zbek tilida polisemiyaga boy fe'llar va sifatlar uchun muhim [4]. Shunga qaramay, kontekstual model ham sinonimlikni avtomatik "tenglashtirish"ga moyil bo'lishi mumkin; chunki u ko'pincha bir xil semantik maydondagi birliklarni yaqin joylashtiradi, lekin uslubiy farq, registr, ijtimoiy-pragmatik cheklovlarni yetarlicha ajrata olmaydi. Demak, mexanizmida sinonimlikni aniqlashni faqat embedding o'xshashligiga bog'lab qo'yish metodik xatodir; u didaktik

jihatdan noto'g'ri tavsiyalar beradi, masalan, neytral va kitobiy birliklarni bir xil vaziyatga tavsiya qilishi mumkin.

Uchinchi bosqichda sinonimlikni baholash mezonlari kiritiladi va bu tezisning asosiy ilmiy urg'usini tashkil etadi. Taklif etilayotgan mexanizm sinonimlikni uch o'lchamda tekshiradi: semantik yadro yaqinligi, kollokatsion moslik va uslubiy-registr moslik. Semantik yadro yaqinligi embeddinglar va leksik ta'riflar uyg'unligida baholanadi; kollokatsion moslik esa nomzod so'zlarning tipik birikmalari, boshqaruv modeli va sintaktik roli orqali tekshiriladi; uslubiy moslik uchun janrlar bo'yicha taqsimot, rasmiy va norasmiy matnlarda uchrash chastotasi, shuningdek, o'quv materiallarida qabul qilingan normativlik indikatorlari ishlatiladi. Bu mezonlar birgalikda qo'llanganda sinonim nomzodlar orasida didaktik jihatdan foydali juftliklar ajralib chiqadi, chunki o'quvchi uchun aynan "qaysi vaziyatda qaysi variant" savoli hal qiluvchi hisoblanadi. Mazkur yondashuvda baholash ko'p mezonli bo'lgani sababli yakuniy qaror skorlash va chegaraviy qiymatlar orqali beriladi; chegaralar o'quv darajasi bilan moslab sozlanadi, ya'ni boshlang'ich bosqichda faqat yuqori darajada almashinadigan sinonimlar, yuqori bosqichlarda esa uslubiy jihatdan farqlanuvchi variantlar ham kiritiladi [1; 5].

To'rtinchi bosqich nazoratli o'qitish konturini yaratishdan iborat bo'lib, bunda "sinonim" va "sinonim emas" belgilari bilan kichik, ammo sifatli belgilangan to'plam tuziladi. O'zbek tili uchun bunday belgilashda mutaxassis lingvist va metodist hamkorligi zarur: lingvist semantik qamrov va normativlikni tekshiradi, metodist esa o'quv maqsadiga moslikni belgilaydi. Model kirish xususiyatlari sifatida embedding juftligi, kollokatsion ko'rsatkichlar, janr taqsimoti va morfosintaktik moslik signallarini oladi; chiqishda esa sinonimlik ehtimoli va didaktik "tavsiya darajasi"ni beradi. Bu yerda muhim jihat shundaki, "tavsiya darajasi" sof lingvistik hukm emas, balki o'qitish sharoitiga bog'liq qarordir: masalan, ilmiy uslub uchun mos sinonim umumiy so'zlashuv mashqida tavsiya etilmasligi mumkin. Natijada NLP modelining chiqishi pedagogik qaror qabul qilishga xizmat qiladigan izohlanadigan ko'rsatkichlarga aylantiriladi, bu esa o'qituvchi va o'quvchi uchun shaffoflikni oshiradi.

Beshinchi bosqich didaktik ssenariylarni avtomatlashtirishni qamrab oladi. Sinonimlarni o'qitishda mashqlar turi faqat "almashtirib qo'yish" bilan cheklanmaydi; sinonimni tanlash vazifalari, uslubiy moslikni topish, gapning pragmatik maqsadiga ko'ra variantni aniqlash, kollokatsion cheklovlarni ko'rish kabi topshiriqlar samaraliroq hisoblanadi. Taklif etilayotgan mexanizm sinonimlar klasteri va ularning registr belgilariga tayangan holda adaptiv mashq generatorini ko'zda tutadi: o'quvchining xatolari qaysi o'lchamda ekanini aniqlab, keyingi topshiriqlarni shu bo'shliqni to'ldirishga yo'naltiradi. Masalan, o'quvchi semantik yadro jihatidan to'g'ri tanlaydi, ammo kollokatsiyada adashsa, tizim birikmalarni solishtirishga qaratilgan topshiriqlarni ko'paytiradi; agar uslubiy moslikda xato bo'lsa, janrlararo taqqoslash mashqlari beriladi. Bu yondashuv sinonimiyani "yodlash" emas, balki kontekstga mos "tanlash kompetensiyasi" sifatida shakllantirishga xizmat qiladi.

Oltinchi bosqich baholash va validatsiya masalalarini yoritadi. NLP tizimida sinonim aniqlashning an'anaviy metrikalari (aniqlik, to'liqlik) didaktik samaradorlikni to'liq ifodalamaydi, chunki ta'limda "xato tavsiya"ning narxi yuqori: o'quvchi noto'g'ri variantni normaga aylantirib olishi mumkin. Shu sababli baholash ikki konturda olib boriladi: birinchi konturda lingvistik sifat metrikalari bilan modelning sinonim juftliklarni ajratish qobiliyati tekshiriladi; ikkinchi konturda esa pedagogik tajriba orqali o'quv natijalari, xatolar spektri va transfer (yangi kontekstda qo'llay

olish) ko'rsatkichlari tahlil qilinadi. Muhimi, sinonim klasterlari bo'yicha "qisman to'g'ri" javoblarni ham baholash konsepsiyasi ishlab chiqiladi: agar o'quvchi semantik yadroga mos, ammo registr jihatdan noaniq variantni tanlasa, tizim uni mutlaq xato deb belgilamaydi, balki izohli teskari aloqani beradi.

NLP sistemasi orqali sinonimlarni o'qitish mexanizmi korpus tayyorlash, sinonim nomzodlarni ajratish, ko'p mezonli baholash, nazoratli o'qitish, adaptiv mashq yaratish va ikki konturli validatsiya bosqichlari birligi sifatida asoslandi. Taklif etilgan yondashuv sinonimlikni faqat semantik o'xshashlik emas, balki kollokatsion va uslubiy moslik bilan birga didaktik tavsiya darajasida talqin qilish imkonini beradi. Natijada o'quvchi sinonimlarni yodlashdan ko'ra, kontekstga mos tanlash va asoslash kompetensiyasini egallashga yo'naltiriladi, o'qituvchi esa tizim tavsiyalarini izohlanadigan ko'rsatkichlar orqali nazorat qilish imkoniga ega bo'ladi.

Foydalanilgan adabiyotlar ro'yxati

1. Rahmatullayev Sh. O'zbek tilining izohli frazeologik lug'ati. Toshkent, O'qituvchi, 1992. 384 b.
2. Yo'ldoshev M. Matn lingvistikasi. Toshkent, Tafakkur, 2008. 216 b.
3. Jurafsky D., Martin J. H. Speech and Language Processing. Boston, Pearson, 2009. 1024 p.
4. Mikolov T., Sutskever I., Chen K., Corrado G. S., Dean J. Distributed Representations of Words and Phrases and their Compositionality. Red Hook, Curran Associates, 2013. P. 3111–3119.
5. Apresyan Yu. D. Leksicheskaya semantika: sinonimicheskie sredstva yazyka. Moskva, Nauka, 1974. 367 s.
6. Vygotskiy L. S. Myshlenie i rech. Moskva, Pedagogika, 1982. 287 s.