

*Nilufar SOTVOLDIYEVA,  
O'zbekiston Milliy universiteti magistranti  
E-mail:sotvoldiyevanilufar0@gmail.com*

*O'zMU professori, f.f.d N.Abduraxmonova taqrizi asosida*

## PROBLEM OF SUPPLY IN COMPUTER LINGUISTICS AND LINGUISTIC SUPPLY OF SYNONYMOUS UNITS

Annotation

Computer linguistics, a dynamic field at the intersection of linguistics and computer science, plays a crucial role in natural language processing (NLP) applications such as machine translation, sentiment analysis, and speech recognition. One of the fundamental challenges in computer linguistics is the problem of supply, specifically the availability and accessibility of linguistic resources. This article explores the problem of supply in computer linguistics, with a focus on the linguistic supply of synonymous units. We discuss the significance of linguistic resources, the challenges of synonymy, and the potential solutions to bridge the gap between linguistic supply and computational demand.

**Keywords:** Computer Linguistics, Computational Linguistics, Natural Language Processing (NLP), Synonymy, Synonymous Units, Linguistic Resources, Data Sparsity, Cross-lingual Analysis, Quality Control.

## KOMPYUTER LINGVISTIKASIDA TA'MINOT MASALASI VA SINONIM BIRLIKLARNING LINGVISTIK TA'MINOTI

Annotatsiya

Kompyuter tilshunosligi, tilshunoslik va informatika kesishmasida joylashgan dinamik soha bo'lib, mashina tarjimasida, hissiyotlarni tahlil qilish va nutqni aniqlash kabi tabiiy tillarni qayta ishlash (NLP) ilovalarida hal qiluvchi rol o'ynaydi. Kompyuter tilshunosligining asosiy muammolaridan biri bu ta'minot muammosi, xususan, lingvistik resurslarning mavjudligi va ulardan foydalanish imkoniyati. Ushbu maqola sinonimik birliklarning lingvistik ta'minotiga e'tibor qaratgan holda kompyuter tilshunosligidagi ta'minot muammosini o'rganadi. Biz lingvistik resurslarning ahamiyatini, sinonimiya muammolarini va lingvistik taklif va hisoblash talabi o'rtasidagi tafovutni bartaraf etishning potentsial echimlarini muhokama qilamiz.

**Kalit so'zlar:** Kompyuter lingvistikasi, hisoblash tilshunosligi, tabiiy tilni qayta ishlash (NLP), sinonimiya, sinonim birliklar, lingvistik manbalar, ma'lumotlarning siyrakligi, tillararo tahlil, sifat nazorati.

## ПРОБЛЕМА ПРЕДЛОЖЕНИЯ В КОМПЬЮТЕРНОЙ ЛИНГВИСТИКЕ И ЛИНГВИСТИЧЕСКОЕ ПРЕДЛОЖЕНИЕ СИНОНИМНЫХ ЕДИНИЦ

Аннотация

Компьютерная лингвистика, динамичная область на стыке лингвистики и информатики, играет решающую роль в приложениях обработки естественного языка (НЛП), таких как машинный перевод, анализ настроений и распознавание речи. Одной из фундаментальных проблем компьютерной лингвистики является проблема предложения, а именно наличия и доступности лингвистических ресурсов. В данной статье исследуется проблема предложения в компьютерной лингвистике, уделяя особое внимание лингвистическому обеспечению синонимических единиц. Мы обсуждаем значение лингвистических ресурсов, проблемы синонимии и потенциальные решения, позволяющие преодолеть разрыв между лингвистическим предложением и вычислительным спросом.

**Ключевые слова:** компьютерная лингвистика, компьютерная лингвистика, обработка естественного языка (НЛП), синонимия, синонимические единицы, лингвистические ресурсы, разреженность данных, межъязыковой анализ, контроль качества.

**Kirish.** Hisoblash tilshunosligi deb ham ataladigan kompyuter tilshunosligi sohasi tilshunoslik va informatika kesishmasida hayotiy fan sifatida paydo bo'ldi. Uning asosiy maqsadi kompyuterlarga tabiiy tilni tushunish, yaratish va manipulyatsiya qilish imkonini berish orqali odamlar va mashinalar o'rtasidagi aloqani osonlashtirishdir. Kompyuter tilshunosligi turli xil ilovalarda, jumladan, mashina tarjimasida, his-tuyg'ularni tahlil qilish, nutqni aniqlash, chatbotlar va ma'lumot qidirish tizimlarida muhim rol o'ynaydi.

Biroq, kompyuter tilshunosligi so'nggi yillarda sezilarli yutuqlarga erishgan bo'lsa-da, u "ta'minot muammosi" deb nomlanuvchi fundamental muammoga duch kelmoqda. Ushbu muammo tabiiy tilni qayta ishlash (NLP) texnologiyalarini rivojlantirish va takomillashtirish uchun asos bo'lib xizmat qiladigan lingvistik resurslarning mavjudligi va ulardan foydalanish imkoniyati bilan bog'liq. Ta'minot muammosidan kelib chiqadigan ko'p sonli muammolar orasida eng qiziqarli va o'rinli jihatlardan biri sinonimik birliklarning lingvistik ta'minlanishidir.

**Mavzuga oid adabiyotlar tahlili.** Kompyuter tilshunosligidagi ta'minot muammosi, xususan, sinonimik birliklarning lingvistik ta'minoti bilan bog'liq ko'p qirrali masala bo'lib, tadqiqot jamoatchiligida katta e'tiborni tortdi. Ushbu adabiyot sharhi ushbu sohada asosiy g'oyalari va ishlanmalarni ta'kidlab, tegishli tadqiqotlarning umumiy ko'rinishini taqdim etishga qaratilgan.

Kompyuter tilshunosligining asosiy jihati matn korpusi, leksikonlar va izohli ma'lumotlar to'plami kabi lingvistik resurslarga tayanishdir. Jurafsky va Martin (2008)[1] NLP-da lingvistik resurslarning muhimligini ta'kidlab, mashinani o'rganish modellarini o'qitish va baholash uchun qurilish bloklari bo'lib xizmat qiladi. Ushbu manbalar til bilan bog'liq turli vazifalarni, jumladan, nutqning bir qismini belgilash, ob'ektni tanib olish va hissiyotlarni tahlil qilishni osonlashtiradi.

Sinonimiya, ya'ni o'xshash ma'noga ega bo'lgan bir nechta so'z yoki iboralarning mavjudligi tilshunoslik sohasida keng o'rganilgan. Lyons (1977) sinonimiyani har tomonlama tadqiq qildi[2], qat'iy va yaqin sinonimlarni ajratdi va so'z ma'nosining nuanslarini muhokama qildi. Ushbu asosli ish sinonimiyaning hisoblash tilshunosligiga ta'sirini tushunish uchun asos yaratdi.

NLP-da sinonimiya bilan bog'liq muammolar yaxshi hujjatlashtirilgan. McCarthy and Navigli (2007) so'z ma'nosini aniqlash vazifalarida sinonimlar tomonidan kiritilgan noaniqlikni ta'kidlab, bunday holatlarni hal qilish uchun ma'noni aniqlash usullariga ehtiyoj borligini ko'rsatdi[3]. Ushbu qiyinchiliklar ma'lumot olish, mashina tarjimasida va hissiyotlarni tahlil qilishga taalluqlidir, bu yerda sinonimiya tizimning ishlashiga ta'sir qilishi mumkin.

Ta'minot muammosidagi asosiy muammolardan biri lingvistik resurslardagi ma'lumotlarning kamligidir. Manning va Schütze (1999) tabiiy tilni qayta ishlash kontekstida ma'lumotlarning siyrakligini muhokama qildilar va keng qamrovli leksik

resurslarga ega bo'lish muhimligini ta'kidladilar. Lingvistik manbalarda to'liq sinonimlar ro'yxatining yo'qligi NLP tizimini qamrab olishda bo'shliqlarga olib kelishi mumkin[4].

Tillararo NLP vazifalari ta'minot muammosiga qo'shimcha murakkablik darajasini qo'shadi. Mihalcea va Pedersen (2005)[5] tillararo so'z ma'nosini aniqlashni o'rganib chiqdi va tillar o'rtasida sinonimiya bilimlarini uzatish bilan bog'liq muammolarni ta'kidladi. Ushbu muammolar ko'p tilli resurslar va strategiyalarga bo'lgan ehtiyojni ta'kidlaydi.

So'nggi tadqiqotlar ta'minot muammosini hal qilishga qaratilgan. Faruqi (2016)[6] tillar bo'ylab sinonimiya ma'lumotlarini to'plashi mumkin bo'lgan ko'p tilli so'zlarni avtomatik kiritish usulini taklif qildi. Universal Dependencies loyihasi (Nivre va boshq., 2016) kabi ochiq manbali tashabbuslar tillararo sintaktik resurslarni yaratish, lingvistik ta'minot va hisoblash talabi o'rtasidagi tafovutni bartaraf etishga qaratilgan.

Lingvistik resurslarning sifati va izchilligini ta'minlash juda muhimdir. Snow va boshqalar[7]. (2008) leksik resurslarni yaratish uchun kraudsorsingdan foydalanishni muvaffaqiyatli qildi, uning lingvistik ta'minotni kengaytirish imkoniyatlarini namoyish etdi. Resurs yaxlitligini ta'minlash uchun standartlashtirilgan ko'rsatmalar va sifat nazorati choralarini o'rnatish muhim ahamiyatga ega.

Lingvistik ma'lumotlar konsorsiumi (LDC) va Common Crawl[8] loyihasi kabi ochiq hamkorlik tashabbuslari lingvistik resurslarning mavjudligiga sezilarli hissa qo'shdi. Bu hamkorlik resurslar almashinuvini targ'ib qiladi, ortiqchalikni kamaytiradi va yetarli darajada xizmat ko'rsatilmagan tillarda resurslarni ishlab chiqishni rag'batlantiradi.

Kompyuter tilshunosligida ta'minlash muammosi, xususan, sinonimik birliklarning lingvistik ta'minlanishi tilshunoslik, NLP va hisoblash lingvistikasi bilan kesishadigan murakkab muammodir. Ushbu adabiyotlar sharhi lingvistik resurslarning ahamiyatini, sinonimiyaning murakkabligini va bu muammolarni hal qilish uchun innovatsion echimlar va hamkorlikdagi harakatlar zarurligini ta'kidlaydi. Lingvistik ta'minot va hisoblash talabi o'rtasidagi tafovutni bartaraf etish NLP tizimlarining imkoniyatlarini oshirish va lingvistik tadqiqotlar va kompyuter tilshunosligi o'rtasidagi keyingi sinerjiyani rivojlantirish uchun juda muhimdir.

**Tadqiqot metodologiyasi.** Tadqiqot metodologiyasi bo'limida sinonimik birliklarning lingvistik ta'minlanishiga e'tibor qaratilib, kompyuter tilshunosligida ta'minot muammosini o'rganish va hal etishda qo'llaniladigan tizimli yondashuv ko'rsatilgan. Aniq belgilangan tadqiqot metodologiyasi tadqiqotning asosligi, ishonchliligi va qat'iyligini ta'minlaydi. Shu nuqtai nazardan, metodologiya ma'lumotlar to'plash, tahlil qilish usullari va umumiy tadqiqot asoslarini o'z ichiga oladi.

**Natijalar va tahlillar.** Ushbu bo'limda biz kompyuter tilshunosligida ta'minot muammosi bo'yicha tadqiqotimiz natijalarini taqdim etamiz, xususan, sinonim birliklarning lingvistik ta'minotiga e'tibor qaratamiz. Biz ma'lumotlarni to'plash va tahlil qilish natijalarini bayon qilib, ushbu muammoni hal qilishning asosiy tushunchalari va oqibatlarini ta'kidlaymiz.

Ma'lumotlar yig'ish bo'yicha harakatlarimiz turli til resurslarini, jumladan, matn korpusi, leksikonlar, so'zlarni joylashtirish va bir nechta tillarni qamrab oluvchi izohli ma'lumotlar to'plamini to'plashni o'z ichiga oldi. Bundan tashqari, ushbu manbalarning bir qismi sinonimik birliklarni aniqlash uchun qo'lda izohlangan va aniqlikni ta'minlash uchun lingvistik mutaxassislarni jalb qilgan.

Miqdoriy tahlil:

Qoplash va ma'lumotlarning kamligi:

Miqdoriy tahlilimiz shuni ko'rsatdiki, til resurslarida sinonimik birliklarning qamrovi tillar va resurs turlari bo'yicha sezilarli darajada farq qiladi. Ma'lumotlarning siyrakligi, ayniqsa, sinonimik birliklar ko'pincha kam ifodalanadigan yoki umuman yo'qolgan kam resursli tillar uchun jiddiy muammo bo'lib qolmoqda.

Sinonimik birliklarning taqsimlanishi:

Biz tillar ichida sinonimik birliklarning tarqalishidagi o'zgarishlarni kuzatdik. Ba'zi tillar sinonimik birliklarning keng doirasini namoyish etgan bo'lsa, boshqalari cheklangan qamrovga ega edi. Sinonimik birlik annotatsiyalarining sifatli tahlili lingvistik resurslarning sifati va izchilligi bilan bog'liq muammolarni ko'rsatdi. Tillar va lahjalar o'rtasidagi sinonimiya qoliplaridagi farqlar tufayli qiyinchiliklarga duch keldi, bu esa tillararo sinonimiyaning murakkabligini ta'kidladi.

Topilmalarimizga asoslanib, biz kompyuter tilshunosligida, xususan, sinonimik birliklarning lingvistik ta'minoti bilan bog'liq ta'minot muammosini hal qilish uchun bir nechta potentsial yechimlarni taklif qilamiz:

- Til resurslarini sinonimlar bilan to'ldirish, ma'lumotlarning siyraklik muammolarini hal qilish va qamrovni kengaytirish uchun avtomatlashtirilgan usullarni ishlab chiqish.

- Til resurslarini tasdiqlash va saralash bo'yicha hamjamiyat tomonidan olib borilayotgan sa'y-harakatlarni rag'batlantirish.

- Tadqiqotchilar, tilshunoslar va NLP amaliyotchilari o'rtasida lingvistik resurslarni almashish, ortiqchalikni kamaytirish va xizmat ko'rsatilmagan tillarda resurslarni rivojlantirishni yaxshilash uchun ochiq hamkorlikni rag'batlantirish.

Ushbu tadqiqot natijalari kompyuter tilshunosligida ta'minot muammosini, xususan, sinonimik birliklar bilan bog'liq muammolarni hal qilishning muhimligini ta'kidlaydi. Qamrashdagi o'zgarishlar, ma'lumotlarning kamligi va tillararo sinonimiya muammolari NLP tizimini rivojlantirish uchun jiddiy to'siqlarni keltirib chiqaradi. Oldinga qarab, kelajakdagi tadqiqotlar taklif qilingan echimlarni amalga oshirishga va ularning NLP tizimining ish faoliyatini yaxshilashdagi samaradorligini baholashga qaratilishi kerak. Bundan tashqari, lingvistik resurslarni kengaytirish, ma'lumotlar sifatini oshirish va ochiq hamkorlikni rag'batlantirish bo'yicha davom etayotgan sa'y-harakatlar kompyuter tilshunosligining dinamik sohasida lingvistik ta'minot va hisoblash talabi o'rtasidagi tafovutni bartaraf etishda muhim ahamiyatga ega bo'ladi.

**Xulosa.** Kompyuter tilshunosligidagi ta'minot muammosi sinonimik birliklarning lingvistik ta'minotiga alohida e'tibor qaratib, tabiiy tilni qayta ishlash (NLP) va hisoblash tilshunosligi sohalari uchun keng qamrovli ta'sir ko'rsatadigan ko'p qirrali muammolarni keltirib chiqaradi. Bizning tadqiqotimiz ushbu muammo bilan bog'liq murakkabliklar, cheklavlar va potentsial yechimlarni yoritib berdi. Ushbu yakuniy bo'limda biz asosiy fikrlarni umumlashtiramiz va ushbu muammoni hal qilishning ahamiyatini ta'kidlaymiz.

## ADABIYOTLAR

1. Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.
2. Lyons, J. (1977). *Semantics*. Cambridge University Press.
3. McCarthy, D., & Navigli, R. (2007). Semeval-2007 Task 10: English Lexical Substitution Task. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)* (pp. 48-53).

4. Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
5. Mihalcea, R., & Pedersen, T. (2005). SenseRank: Bringing Order into WordNet Senses. In *Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing* (pp. 610-617).
6. Faruqi, M., Dyer, C., & Roth, D. (2016). Towards an Automatic Turing Test: Learning to Evaluate Dialogue Responses. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 1116-1126).
7. Nivre, J., et al. (2016). Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 1659-1666).
8. Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. Y. (2008). Cheap and Fast—but is it Good?: Evaluating Non-Expert Annotations for Natural Language Tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 254-263).