

ISSUES OF USING ONTOLOGY IN LINGUISTIC CORPORA

Mamarajabov Abduvali Yuldosh o'g'li

Lecturer, Department of Uzbek Language and Literature Termiz State University

Abstract. This scientific work provides an in-depth analysis of the theoretical and practical foundations of using ontology in linguistic corpora. Ontology is considered an important methodological tool for semantically modeling corpus data, identifying conceptual relationships between linguistic units, and forming knowledge bases.

The study highlights the integration of the ontological approach with semantic annotation, information retrieval systems, and artificial intelligence-based NLP models. In addition, the prospects and existing challenges of creating an ontological corpus in Uzbek linguistics are scientifically analyzed.

Keywords: linguistic corpus, ontology, semantic modeling, knowledge base, corpus linguistics, NLP, conceptual system, semantic network, data integration, automatic analysis

LINGVISTIK KORPUSLARDA ONTOLOGIYADAN FOYDALANISH MASALALARI

Mamarajabov Abduvali Yuldosh o'g'li

TerDU O'zbek tili va adabiyoti kafedrası o'qituvchisi

Annotatsiya. Ushbu ilmiy ishda lingvistik korpuslarda ontologiyadan foydalanishning nazariy va amaliy asoslari chuqur tahlil qilinadi. Ontologiya korpus ma'lumotlarini semantik jihatdan modellashtirish, lingvistik birliklar o'rtasidagi konseptual munosabatlarni aniqlash va bilimlar bazasini shakllantirishda muhim metodologik vosita sifatida qaraladi. Tadqiqotda ontologik yondashuvning semantik annotatsiya, axborot qidiruv tizimlari va sun'iy intellekt asosidagi NLP modellar bilan integratsiyasi yoritiladi. Shuningdek, o'zbek tilshunosligida ontologik korpus yaratish istiqbollari va mavjud muammolar ilmiy asosda tahlil qilinadi.

Kalit soʻzlar: lingvistik korpus, ontologiya, semantik modellashtirish, bilimlar bazasi, korpus lingvistikasi, NLP, konseptual tizim, semantik tarmoq, maʼlumotlar integratsiyasi, avtomatik tahlil

Kirish. Zamonaviy tilshunoslikda lingvistik korpuslar til birliklarini real kommunikativ muhitda oʻrganish imkonini beruvchi eng muhim ilmiy resurslardan biri hisoblanadi. Biroq korpus lingvistikasining rivojlanishi natijasida faqat statistik va strukturaviy tahlil yetarli emasligi ayon boʻlib bormoqda. Tilning chuqur semantik qatlamini ochib berish uchun konseptual va formal modellashtirish zarurati paydo boʻlmoqda. Shu nuqtai nazardan ontologiya lingvistik korpuslarni yangi bosqichga olib chiquvchi asosiy metodologik yondashuv sifatida maydonga chiqadi. Ontologiya yordamida lingvistik birliklar oʻrtasidagi mantiqiy va semantik munosabatlar formal tizimga keltiriladi. Bu esa korpus maʼlumotlarini nafaqat saqlash, balki ularni bilimlar tizimi sifatida qayta ishlash imkonini beradi. Natijada korpus lingvistikasi sunʼiy intellekt, bilimlar muhandisligi va kompyuter lingvistikasi bilan integratsiyalashgan interdisiplinar sohaga aylanadi.

Asosiy qism. Lingvistik korpuslarda ontologiyadan foydalanish zamonaviy kompyuter lingvistikasi va raqamli tilshunoslikning eng muhim yoʻnalishlaridan biri hisoblanadi. Ushbu yondashuv til birliklarini faqat statistik yoki formal jihatdan emas, balki semantik va konseptual darajada ham tahlil qilish imkonini beradi. Ontologiya korpus maʼlumotlarini tizimlashtirish, lingvistik birliklar oʻrtasidagi mantiqiy bogʻlanishlarni aniqlash hamda bilimlar bazasini yaratishda asosiy metodologik vosita sifatida xizmat qiladi [4, 12-b.].

Lingvistik korpuslar dastlab matnlarning katta hajmdagi toʻplami sifatida qaralgan boʻlsa, hozirgi kunda ular bilimlar tizimi sifatida talqin qilinmoqda. Bu oʻzgarish ontologik yondashuvning kirib kelishi bilan bogʻliq boʻlib, u korpusdagi har bir lingvistik birlikni konseptual obyekt sifatida modellashtirish imkonini beradi. Natijada

soʻzlar, soʻz birikmalari va gaplar oʻzaro bogʻlangan semantik tarmoq elementlariga aylantiriladi [2, 15-b.]. N. Abduraxmonova tomonidan ishlab chiqilgan elektron korpus modeli oʻzbek tilining formal-funksional xususiyatlarini hisobga olgan holda yaratilgan boʻlib, unda lingvistik birliklar nafaqat morfologik, balki semantik jihatdan ham tizimlashtirilgan. Ushbu modelda ontologik yondashuv korpusni bilimlar bazasiga aylantirishning asosiy vositasi sifatida koʻriladi [1, 9-b.]. Bu esa oʻzbek tilining raqamli resurslarini rivojlantirishda muhim ilmiy asos boʻlib xizmat qiladi.

Ontologiyaning lingvistik korpuslardagi asosiy vazifalaridan biri semantik munosabatlarni formal tarzda ifodalashdir. Bu jarayonda “is-a”, “part-of”, “related-to” kabi mantiqiy bogʻlanishlar orqali lingvistik birliklar oʻzaro bogʻlanadi. Shu orqali korpus maʼlumotlari oddiy matnlar yigʻindisidan bilimlar grafigiga aylanadi. Bunday grafik model zamonaviy sunʼiy intellekt tizimlarida keng qoʻllaniladi va u knowledge graph sifatida tanilgan [5, 22-b.]. C. Chiarcos tomonidan ishlab chiqilgan lingvistik annotatsiyalar ontologiyasi til birliklarini formal kategoriyalarga ajratish va ularning oʻzaro munosabatlarini aniqlashga xizmat qiladi. Ushbu yondashuvga koʻra, har bir lingvistik birlik mustaqil element emas, balki maʼlum bir semantik tizimning tuguni sifatida qaraladi [4, 18-b.]. Bu esa korpus lingvistikasida yangi konseptual model yaratish imkonini beradi.

Ontologik modellashtirish jarayonida lingvistik maʼlumotlar bosqichma-bosqich formal tizimga keltiriladi. Avval lingvistik birliklar aniqlanadi, soʻng ular konseptual kategoriyalarga ajratiladi va nihoyat ular oʻrtasidagi semantik munosabatlar belgilanadi. C. M. Keetning tadqiqotlariga koʻra, ontologiya bilimlarni kompyuter tizimlari tomonidan qayta ishlashga moslashtiruvchi eng muhim vositalardan biridir [5, 17-b.]. Lingvistik korpuslarda ontologiyaning yana bir muhim jihati semantik noaniqlikni kamaytirishdir. Tabiiy tilning koʻp maʼnoliligi (polysemy) va omonimiya hodisalari korpus tahlilini murakkablashtiradi. Ontologik yondashuv esa har bir maʼno uchun

alohida konseptual tugun yaratish orqali bu muammoni hal qiladi. Masalan, bir xil shakldagi soʻzlar turli ontologik kategoriyalarga ajratiladi va ularning kontekstual maʼnosi aniq belgilanadi [3, 21-b.]. M.Abjalova tadqiqotlarida taʼkidlanishicha, korpus lingvistikasida ontologik yondashuv semantik annotatsiya jarayonini sezilarli darajada takomillashtiradi, chunki u lingvistik birliklarni faqat belgilash emas, balki ularning oʻzaro semantik aloqalarini ham tizimlashtiradi [2, 20-b.]. Bu esa korpusni oddiy matnlar bazasidan bilimlar tizimiga aylantiradi.

Shuningdek, Arabic Ontology Project doirasida ishlab chiqilgan model ham lingvistik ontologiyaning amaliy imkoniyatlarini kengaytiradi. Ushbu modelda WordNet asosidagi konseptlar tizimi qoʻllanilib, til birliklari oʻrtasidagi semantik bogʻlanishlar avtomatik ravishda aniqlanadi [6, 14-b.]. Bu yondashuv koʻp tilli korpuslarni yaratishda muhim ahamiyatga ega. Ontologik korpuslar axborot qidiruv tizimlarida ham katta ahamiyat kasb etadi. Anʼanaviy qidiruv tizimlari faqat kalit soʻzlarga asoslangan boʻlsa, ontologik tizimlar maʼno asosida ishlaydi. Bu esa foydalanuvchi soʻrovining kontekstini hisobga olish imkonini beradi va natijada aniqroq axborot olishga yordam beradi. Biroq ontologik modellashtirish jarayoni murakkab va resurs talab qiluvchi jarayon boʻlib, katta hajmdagi lingvistik maʼlumotlarni standartlashtirish va doimiy yangilab borishni talab qiladi. Bundan tashqari, avtomatik ontologiya yaratishda semantik xatoliklar ehtimoli mavjud boʻlib, bu inson ekspertizasi bilan tekshirish zaruratini keltirib chiqaradi. Lingvistik korpuslarda ontologiyadan foydalanish zamonaviy tilshunoslikni yangi bosqichga olib chiqmoqda. U tilni faqat kommunikativ vosita sifatida emas, balki formal bilimlar tizimi sifatida oʻrganish imkonini beradi va lingvistika, informatika hamda sunʼiy intellekt oʻrtasidagi integratsiyani kuchaytiradi.

Lingvistik korpuslarda ontologiyadan foydalanish jarayoni bugungi kunda faqat nazariy model sifatida emas, balki amaliy raqamli lingvistikaning asosiy texnologik

yoʻnalishi sifatida ham rivojlanmoqda. Ontologik yondashuvning kengayishi korpus maʼlumotlarini qayta ishlashda yangi imkoniyatlarni ochib berdi, xususan, semantik integratsiya, avtomatik tahlil va bilimlar grafigini yaratish jarayonlarini sezilarli darajada takomillashtirdi. Bu jarayonda korpusdagi har bir lingvistik birlik mustaqil element sifatida emas, balki keng konseptual tizimning bir boʻlagi sifatida qaraladi. Natijada til birliklari oʻrtasidagi munosabatlar faqat sintaktik yoki leksik darajada emas, balki chuqur semantik darajada ham ifodalanadi [4, 20-b.].

Ontologiyaning eng muhim xususiyatlaridan biri uning formalizatsiya imkoniyatidir. Formalizatsiya jarayoni til birliklarini aniq matematik va mantiqiy struktura asosida ifodalashni nazarda tutadi. Bu esa korpus maʼlumotlarini kompyuter tizimlari uchun tushunarli va qayta ishlanadigan shaklga keltiradi. Aynan shu jihat ontologiyani oddiy lingvistik tasniflashdan farqlaydi. Chunki anʼanaviy korpus lingvistikasida maʼlumotlar asosan statistik asosda tahlil qilinsa, ontologik yondashuvda maʼno va tushuncha markaziy oʻringa chiqadi [5, 19-b.].

Abjalova M. tomonidan olib borilgan tadqiqotlarda ontologiyaning korpus lingvistikasidagi oʻrni semantik tizimni shakllantirish bilan bevosita bogʻliq ekani taʼkidlanadi. Uning fikricha, ontologiya til birliklarini faqat belgilash vositasi emas, balki ularni tizimli bilimlar strukturasiga aylantiruvchi asosiy mexanizmdir [2, 23-b.]. Bu yondashuv natijasida korpuslar endi faqat matnlar yigʻindisi emas, balki bilimlar bazasi sifatida qaraladi.

Ontologik korpuslarda eng muhim jarayonlardan biri semantik inferensiya, yaʼni mavjud maʼlumotlardan yangi xulosalar chiqarish imkoniyatidir. Masalan, agar korpusda “oʻqituvchi – taʼlim jarayonida ishtirok etadi” va “taʼlim jarayoni – bilim uzatish tizimi” kabi munosabatlar mavjud boʻlsa, tizim avtomatik ravishda “oʻqituvchi bilim uzatish jarayonining ishtirokchisi” degan yangi semantik xulosa chiqarishi mumkin. Bu esa sunʼiy intellekt tizimlarida avtomatik fikrlash mexanizmlarini

rivojlantiradi. Chiarcos tomonidan ishlab chiqilgan lingvistik annotatsiya ontologiyasi ham aynan shu maqsadga xizmat qiladi. Unda til birliklari o‘rtasidagi munosabatlar formal grafik model sifatida ifodalanadi va har bir annotatsiya ma’lum bir semantik kategoriya bilan bog‘lanadi [4, 17-b.]. Bu esa korpusni nafaqat tahlil qilish, balki uni vizual va konseptual darajada ham tushunish imkonini beradi.

O‘zbek tilshunosligida ham ontologik yondashuv asosida korpuslarni rivojlantirishga bo‘lgan qiziqish ortib bormoqda. Abduraxmonova N. tomonidan yaratilgan elektron korpus modeli bu yo‘nalishdagi muhim ilmiy qadam hisoblanadi. Ushbu modelda til birliklari formal-funksional tamoyillar asosida tizimlashtirilgan bo‘lib, ular semantik jihatdan ham bog‘langan [1, 12-b.]. Bu esa o‘zbek tilining raqamli resurslarini rivojlantirishda muhim asos bo‘lib xizmat qiladi.

Ontologik modellashtirishning yana bir muhim jihati uning ko‘p tillilikni qo‘llab-quvvatlashidir. Zamonaviy korpuslar ko‘pincha multilingual xarakterga ega bo‘lib, bir nechta tillar o‘rtasidagi semantik moslikni aniqlash muhim vazifalardan biridir. Ontologiya yordamida turli tillardagi konseptlar yagona semantik tizimga birlashtiriladi, bu esa tarjima tizimlari va cross-lingual NLP modellarining samaradorligini oshiradi [6, 16-b.]. Shuningdek, ontologik korpuslar axborot qidiruv tizimlarida ham muhim o‘rin tutadi. Semantik qidiruv tizimlari foydalanuvchi so‘rovini faqat kalit so‘zlar asosida emas, balki uning konseptual mazmuni asosida tahlil qiladi. Bu esa an’anaviy qidiruv tizimlariga nisbatan ancha aniq va chuqur natijalar beradi. Natijada foydalanuvchi faqat so‘z emas, balki ma’no asosida axborotga ega bo‘ladi. Biroq ontologik modellashtirish jarayoni bir qator muammolar bilan ham bog‘liq. Eng asosiy muammolardan biri lingvistik ma’lumotlarning standartlashtirilmaganligidir. Turli korpuslar va tadqiqotlar o‘rtasida yagona ontologik modelning yo‘qligi integratsiya jarayonini qiyinlashtiradi. Bundan tashqari, avtomatik ontologiya yaratish jarayonida semantik xatoliklar yuzaga kelishi mumkin, bu esa inson ekspertizasi bilan tekshirish zaruratini keltirib chiqaradi.

Xulosa

Lingvistik korpuslarda ontologiyadan foydalanish zamonaviy tilshunoslikni yangi ilmiy bosqichga olib chiqmoqda. Ontologiya korpus ma'lumotlarini semantik va konseptual jihatdan tizimlashtirib, ularni bilimlar bazasiga aylantiradi. Bu esa til birliklari o'rtasidagi mantiqiy munosabatlarni aniqlash va avtomatik xulosa chiqarish imkonini beradi. Natijada korpus lingvistikasi sun'iy intellekt va NLP tizimlari bilan integratsiyalashgan interdisiplinar sohaga aylanadi. Shuningdek, bu yondashuv o'zbek tilshunosligida raqamli resurslarni rivojlantirish va global ilmiy muhitga integratsiya qilish uchun muhim metodologik asos bo'lib xizmat qiladi.

Foydalanilgan manbalar

1. Nilufar Abdurakhmonova. "FORMAL-FUNCTIONAL MODELS OF THE UZBEK ELECTRON CORPUS", September 26, 2021.
<https://doi.org/10.5281/zenodo.5528953>.
2. Abjalova M. Korpus lingvistikasi. [Matn] : uslubiy qo'llanma / M.A. Abjalova. - Toshkent: Bookmany print, 2022. -103 b.
3. O'zbek tili ontologiyasi: yaratish texnologiyasi va konsepsiyasi [Matn]: M.A. Abjalova. - Toshkent: Nodirabegim, 2022. - 212 b.
4. Chiarcos C. An ontology of linguistic annotations [Elektron resurs]. – Semantic Scholar, 2008. – URL: <https://www.semanticscholar.org/paper/An-ontology-of-linguistic-annotations> (murojaat sanasi: 09.05.2026).
5. Keet C. M. Ontology Engineering for NLP Systems [Elektron resurs]. – arXiv, 2017. – URL: <https://arxiv.org/abs/1709.01991>
6. Arabic Ontology Project. Arabic Ontology based on WordNet [Elektron resurs]. – arXiv, 2022. – URL: <https://arxiv.org/abs/2205.09664>