



UDK: 81.03

Madinabonu QODIROVA,
Toshkent davlat o'zbek tili va adabiyoti universiteti tayanch doktoranti
E-mail: begonammm@gmail.com

NamDU dotsenti U.Qo'ziyev taqrizi asosida

DATABASE OF SYNONYM AUXILIARY WORDS FOR MACHINE TRANSLATION

Annotation

This article analyzes the problems encountered in the process of introducing the synonymy of auxiliary words in the Uzbek language into the linguistic database. In particular, despite the fact that auxiliary words express grammatical meaning, they differ in their location in the sentence and syntactic function, the features of their connection with various grammatical suffixes, the limitations of the semantic field, and the existence of synonyms that are not fully compatible with each other are discussed. The study substantiates the need to clearly define the grammatical and semantic characteristics of auxiliary words when creating a linguistic database.

Key words: Automatic translation, synonyms, auxiliary language units, linguistic database, grammatical form, grammatical meaning.

ПРОБЛЕМЫ СОЗДАНИЯ ЛИНГВИСТИЧЕСКОЙ БАЗЫ ДАННЫХ СИНОНИМОВ ВСПОМОГАТЕЛЬНЫХ СЛОВ УЗБЕКИСТАНА ДЛЯ МАШИННОГО ПЕРЕВОДА

Аннотация

В данной статье анализируются проблемы, возникающие в процессе введения синонимов служебных слов узбекского языка в лингвистическую базу данных. В частности, несмотря на то, что служебные слова выражают грамматическое значение, обсуждаются их различия в положении и синтаксической функции в предложении, особенности их связи с различными грамматическими суффиксами, ограниченность семантического поля, а также существование синонимов, не полностью совместимых друг с другом. Результаты исследования обосновывают необходимость четкого определения грамматических и семантических свойств служебных слов при создании языковой базы.

Ключевые слова: Автоматический перевод, синонимы, вспомогательные языковые единицы, лингвистическая база данных, грамматическая форма, грамматическое значение.

MASHINA TARJIMASI UCHUN O'ZBEK TILI YORDAMCHI SO'ZLARI SINONIMIYASI LINGVISTIK BAZASINI YARATISH MUAMMOLARI

Аннотация

Ushbu maqolada o'zbek tilidagi yordamchi so'zlarning sinonimiyasini lingvistik bazaga kiritish jarayonida uchraydigan muammolar tahlil qilinadi. Jumladan, yordamchi so'zlarning grammatik ma'no ifodalashiga qaramay, gapdagi joylashuvi va sintaktik vazifasiga ko'ra farqlanishi, turli grammatik qo'shimchalar bilan bog'lanish xususiyatlari, semantik maydonning cheklovlari va o'zaro to'liq mos kelmaydigan sinonimlarning mavjudligi muhokama qilinadi. Tadqiqot natijasida lingvistik bazani yaratishda yordamchi so'zlarning grammatik va semantik xususiyatlarini aniq belgilash zarurligi asoslab beriladi.

Kalit so'zlar: Avtomatik tarjima, sinonim so'zlar, yordamchi til birliklari, lingvistik ma'lumotlar bazasi, grammatik shakl, grammatik ma'no.

Kirish. Jahon tilshunosligida XX asrning 50-yillaridan e'tiboran matnning elektron hisoblash mashinalari vositasida tahlil qilish muammosi o'rganila boshladi, natijada zamonaviy axborot texnologiyalari dasturlari yuzaga keldi. Bu asosda yuz bergan texnologik taraqqiyot dunyodagi yetakchi adabiy tillarda matnning kvantitativ tahlil qilish tizimlarini yaratishga olib kelgani ma'lum. Zero, globallashtirish va jadal rivojlanayotgan fan-texnika davrida katta hajmdagi axborot ustida tezlik bilan ishlash va samarali natija olish juda zarur. Shu ma'noda, dunyo tilshunosligi fani oldiga tilni lingvistik modellashtirish, so'zlarning tahlil algoritmini tuzish hamda lingvoanaliz dasturining lingvistik modullarini yaratish vazifasi qo'yilmoqda. O'zbekiston Respublikasida 2003-yilning 11-dekabrida "Axborotlashtirish to'g'risida"[1] gi qonunni qabul qilinishi ham bejiz emas. Ushbu qonunning Axborotlashtirish sohasidagi davlat siyosati 4-moddasida Axborot texnologiyalarining zamonaviy vositalari ishlab chiqarilishini tashkil etish ta'lim sohasida ishlab chiqariluvchi dastur tizimi yoki elektron qo'llanmalarni yaratish lozimligi aytib o'tilgan.

Yordamchi so'zlar til tizimida muhim grammatik birliklar bo'lib, ularning ma'nosi va sintaktik xususiyatlari tarjima jarayonida katta ahamiyat kasb etadi. O'zbek tilidagi yordamchi so'zlarning sinonimiyasini lingvistik bazaga kiritish jarayonida turli murakkabliklar kuzatiladi. Chunki ba'zi yordamchi so'zlar bir xil grammatik ma'noni ifodalasa ham, sintaktik vazifasi va gapdagi joylashuviga qarab o'zaro farqlanadi.

Bundan tashqari, ayrim yordamchi so'zlar mustaqil holda sinonim bo'la olmaydi, balki faqat muayyan grammatik shakllar bilan qo'llanganda ekvivalentlik hosil qiladi. Masalan, sabab ma'nosini ifodalovchi "sababli", "tufayli", "uchun" so'zlari "dan keyin" bilan sinonim bo'lishi uchun maxsus kelishik qo'shimchalari talab qilinadi. Shuningdek, ba'zi yordamchi so'zlar faqat ma'lum kontekstda sinonim bo'lib, boshqa holatlarda o'zaro almashinmaydi.

Mazkur maqolada o'zbek tilidagi yordamchi so'zlarning sinonimiyasini lingvistik bazaga kiritish jarayonida yuzaga keladigan muammolar tahlil qilinadi. Ushbu tadqiqot natijalari mashina tarjimasini avtomatlashtirilgan til ishlov berish tizimlari uchun muhim ahamiyatga ega bo'lib, tarjima aniqligini oshirishga xizmat qiladi.

Adabiyotlar tahlili. Lingvistik tipologiya va tillarning qiyosiy tahliliga oid tadqiqotlar generativ tilshunoslik aspektida ko'rib chiqiladi. Kompyuter bazasi uchun tilni modellashtirish eng qulay o'rganish usuli hisoblanadi. Kompyuter morfologiyasi yuqori darajadagi ilovalar uchun zarur, shuning uchun morfologiya doirasida modellashtirish ko'plab tadqiqotlarda qo'llaniladi [2]. Xususan, Karnegi Mellon universitetida M.Raab "Mashina tarjimasini uchun tilni modellashtirish texnikasi" mavzusidagi nomzodlik dissertatsiyasini himoya qildi. Tadqiqot ishiga tilni modellashtirishning turli usullari, yaratilgan til modellariga sharhlar, tadqiqotchining amaliy tadqiqotlari, natijalar va xulosalar kiritilgan [3]. Hayfa universiteti olimlari Gennadiy

Lemberskiy, Noam Ordan, Shuli Vintner maqsadli tildagi asl matnlardan tuzilgan til modellari va maqsadli tilga qo'lda tarjima qilingan matnlardan tuzilgan til modellari o'rtasidagi farqni o'rganadilar. Shu tarzda ular mashina tarjimasi uchun til modellari yaratishni taklif qiladilar [4]. Turk tillarida ham mashina tarjimasi doirasida tadqiqot ishlari olib borilmoqda. Turk olimlari turkcha o'zaro tillar va turkcha-inglizcha mashina tarjimasi bo'yicha tadqiqotlar olib bormoqda. Masalan, A.C.Tantug [5], Z.Sagay [6], I.Hamzao'g'li [7], K.Oltintosh [8], M.O'rxun kabi ko'plab olimlarni sanab o'tish mumkin. Masalan, M.O'rxunning "Uyg'ur tili bilan turkcha o'rtasida mashina tarjimasi" nomli tadqiqot asarida turkiy va uyg'ur tili birliklarining modellari yaratilib, taqqoslanadi [9].

O'zbek kompyuter tilshunosligida bugungi kunda modellashtirishga katta e'tibor berilmoqda. Chunki u bugungi NLPning ko'plab muammolarini hal etishi tasdiqlangan. Shu sababli tilning turli darajalarini modellashtirish bo'yicha ilmiy tadqiqot ishlari olib borilmoqda. Xususan, bu borada B. Elov [10], N. Abdurahmonova, Sh. Hamroyeva [11], S. Bozorova [12], S. Muhammedova [13], S. Xudayarova [14], M. Suyunova [15] kabilarni sanab o'tish mumkin. Ular o'z ishlarida til birliklarini modellashtirishning turli usullarini taklif qilishadi. Xususan, N.Abdurahmonova tomonidan mashina tarjimasi uchun oddiy o'zbek-ingliz jumllarining modellari ishlab chiqilgan va qiyoslangan. Sh. Hamroyevaning doktorlik dissertatsiyasi o'zbek tili morfoanalizatori uchun grammatik shakllarni modellashtirishga bag'ishlangan. B. Elov esa til korpus matnlarini modellashtirishning yashirin dirix usulini taklif etadi. S.Bozorova erkin kontekst grammatikasiga asoslangan sintaktik tahlil usullari va o'zbekcha so'z birikmalari modellari taklif etadi. M. Suyunova o'zbek tilidagi grammatik omonimlarni, S. Xudayarova esa so'z birikmalarini modellashtirish bilan shug'ullanadi.

Morfologik modellashtirishda to'g'ri va aniq lingvistik tavsif berish muhim ahamiyatga ega. Morfologik mezonlarning to'g'riligi quyidagi bosqichlarda tekshiriladi:

1. faqat to'g'ri tuzilgan grammatik so'z shakllarini tahlil qilish;
2. grammatik jihatdan noto'g'ri bo'lgan so'z shakllari qanday aniqlanadi [11].

So'z shakllarining to'g'riligi faqat to'g'ri deb topilgan modellardan aniqlanadi va qolgan holatlar uning noto'g'riligini tasdiqlaydi.

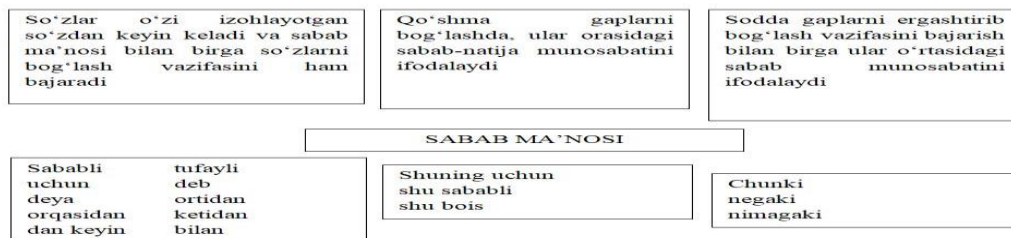
Tabiiy tilni modellashtirish tarjima muammolarining mumkin bo'lgan yechimi hisoblanadi. Xususan, Germaniyaning Axen texnologiya universiteti olimlari Frans Yozef Oh va Hermann Ney Model log-chiziqli modellashtirish usulidan foydalangan holda matn kontekstini hisobga oladigan tarjima usulini taklif qilishdi. Ular, shuningdek, so'z ekvivalentlik sinflari asosida ierarxik leksik modellarni qurishni taklif qilishadi. Bundan tashqari, tadqiqot bog'langan jumllarda so'z tartibini o'rganishga o'rgatish mumkin bo'lgan dasturiy ta'minot tizimini yaratishni taklif qiladi [12].

Tadqiqot metodologiyasi. O'zbek tilida ko'makchi til birliklari sifatida umumlashuvchi bog'lovchi, ko'makchi va yuklamalar o'zbek tilida o'ziga xos shaklga ega bo'lib, ingliz tilida bir necha xil vositalar orqali namoyon bo'ladi. Bunda grammatik shakl bilan grammatik ma'noning mos kelmasligi holatlari kuzatiladi, ya'ni grammatik polisemiya, omonimiya, sinonimiya kabi holatlar yuzaga keladi. Bunday til hodisalari mashina tarjimasida tarjima sifatiga ta'sir qiladi.

Bu yo'nalishda qoidaga asoslangan, ya'ni grammatik tahlilning ahamiyati shundan iboratki, tahlilning ko'p ma'nolilik va omonimiya lisoniy hodisalariga mos kelishi, yana bir jihat har xil turlarga mansub tillardagi turli kategoriyalarni alohida o'rganishdir [13].

Muhokamalar va natijalar. O'zbek tili yordamchi so'zlar sinonimiyasi lingvistik bazasini yaratishda quyidagi muammolar ko'zga tashlanadi:

Bir xil grammatik ma'no ifodalashiga qaramay joylashuv o'rniga ko'ra farqlanuvchi sinonimlar mavjud. Ular lingvistik bazada alohida belgilashni talab etadi. Leksik birliklardan farqli ravishda grammatik birliklarda bir xil grammatik ma'noni ifodalash sinonimlikka asos bo'lmasligi mumkin. Masalan, sabab ma'nosini ifodalovchi yordamchi so'zlar: sababli, tufayli, uchun, deb, deya, orqasidan, ortidan, ketidan, bilan, ...dan keyin, shuning uchun, shu sababli, shu bois, chunki, negaki, nimagaki so'zlari. Ularni gapda tutgan o'rniga ko'ra 3 guruhga bo'lib olamiz. Buni 1-chizmada ko'rishimiz mumkin:



1-chizma. Joylashish o'rniga ko'ra farqlanuvchi yordamchi so'zlar

Demak, ushbu yordamchi so'zlar quyidagi xususiyatiga ko'ra farqlanadi:

Gapdagi vazifasi.

Joylashish o'rnini.

bazada alohida-alohida berilishni talab etadi.

Ayrim grammatik so'zlar o'rnida grammatik so'zga qo'shimcha qo'shibgina o'rindosh sifatida qo'llash mumkin bo'lgan holatlar uchraydi. Misol uchun, sabab ma'nosini ifodalovchi sababli, tufayli, uchun so'zlariga keyin ko'makchisi ham sinonim sifatida qo'llanishi mumkin. Biroq buning uchun keyin so'zi –dan chiqish kelishigi bilan birga qo'llanishi lozim.

Falokat xavfi tug'ilgani sababli ular Yerning Shimoliy qutbiga yo'l olishadi.

Falokat xavfi tug'ilgani tufayli ular Yerning Shimoliy qutbiga yo'l olishadi.

Falokat xavfi tug'ilgani uchun ular Yerning Shimoliy qutbiga yo'l olishadi.

Falokat xavfi tug'ilganidan keyin ular Yerning Shimoliy qutbiga yo'l olishadi.

Ayrim sinonim yordamchi so'zlar o'zidan oldin turli xil grammatik qo'shimchani talab qilib keladi. Masalan, ish-harakatning tezlikda sodir bo'lishini quyidagicha ifodalash mumkin:

Mehmon kirdi-da, allanimani qidira boshladi

Mehmon kiribqoq, allanimani qidira boshladi

Mehmon kirishi bilan, allanimani qidira boshladi.

Ikki obyektini o'zaro qiyoslash uchun ishlatiluvchi ko'makchilar o'zidan oldin ikki xil grammatik shaklni talab etadi.

Senga qaraganda ishni yaxshiroq bajaradi.

Sendan ko'ra ishni yaxshiroq bajaradi.

Xoslik ma'nosini ifodalovchi asosan, binoan, muvofiq, ko'ra so'zlari -ga kelishigi bilan qo'llangandagina bo'yicha so'zi bilan sinonim bo'ladi.

Qonunga asosan ish ko'ring.

Qonunga binoan ish ko'ring

Qonunga muvofiq ish ko'ring

Qonunga ko'ra ish ko'ring

Qonun_ bo'yicha ish ko'ring

Bunday hollarda yordamchi so'zni lingvistik bazada o'zidan oldingi grammatik qo'shimcha bilan birga berish talab etiladi.

Bir grammatik ma'no bir o'rinda bitta grammatik shakl bilan, bir o'rinda 2-3 ta grammatik shakl bilan ifodalanishi mumkin. Misol uchun, garchand...sa ham, sa-da, sa...ham birliklari to'siqsizlik ma'nosini ifodalaydi.

Garchand otasining so'zlari esida bo'lsa ham, qizlarning qistovini rad eta olmadi.

Garchi otasining so'zlari esida bo'lsa ham, qizlarning qistovini rad eta olmadi.

Otasining so'zlari esida bo'lsa ham, qizlarning qistovini rad eta olmadi.

Otasining so'zlari esida bo'lsa-da, qizlarning qistovini rad eta olmadi.

Bunday so'zlar lingvistik bazada alohida belgilanishi lozim.

Bir grammatik ma'noni bildirsa-da, belgini ifodalash darajasi, qo'llanuvchi semantik maydoni bilan bir-biridan farqlanadi, ya'ni biri o'rinda ikkinchisini qo'llab bo'lmaydi.

obyektga nisbatan mustasnoqlik	tashqari	o'zga	boshqa	bo'lak	Fayzulladan tashqari hamma sinfdan chiqsin
ish-harakatga nisbatan mustasnoqlik	bo	bo	o		Yordam chaqirishdan bo'lak choramiz qolmadi
	'lak	shqa	'zga		

1-jadval. Birga kelgan so'zlariga ko'ra farqlanuvchi yordamchi so'zlar.

Shuningdek, ba'zi hollarda yordamchi so'zga birikib kelayotgan so'zning shakli tomonini ham e'tiborga olish zarur bo'ladi:

Darsdan so'ng ko'rishishga kelishdik

Darsdan keyin ko'rishishga kelishdik

Darsdan nari ko'rishishga kelishdik.

So'ng, keyin, nari sinonimik qatoridagi yordamchi so'zlar bir shaklli so'zga ergashib kelganda sinonimlik hosil qilsa-da, so'z birikmasiga bog'lanib kelganda nari ko'makchisi sinonim bo'la olmaydi.

Qo'ng'iroq chalingandan so'ng dars boshlandi.

Qo'ng'iroq chalingandan keyin dars boshlandi.

Xulosa. XX asrning ikkinchi yarmidan boshlab, lingvistik tadqiqotlarda axborot texnologiyalari vositasida matn tahlil qilish muhim ahamiyat kasb eta boshladi. Bu jarayon turli tillarda kvantitativ tahlil usullarini ishlab chiqishga hamda lingvistik modellashirish asosida tilni avtomatlashtirilgan qayta ishlash tizimlarini yaratishga imkon berdi. O'zbek tilshunosligida ham til

Ichida, ustida, bilan, ichra kabi so'zlar ruhiy holatni bildiradi. Lekin doim ham biri o'rinda ikkinchisini qo'llab bo'lmaydi.

Jahl ustida qanday javob berganimni bilmayman.

Xayol ichra suzardi

Har ikki gapdagi yordamchi so'zlarning grammatik ma'nosi ruhiy holat sanaladi va lingvistik bazada shu nom bilan berish kerak bo'ladi. Bu bazada so'zlarni moslashtirishda muammolar keltirib chiqaradi.

So'zlarning bir-biriga aynan mos kelmasligi. Muayyan jumalardagina sinonim bo'lib boshqa o'rinlarda ma'noviy jihatdan mos kelmaslik holatlari mavjud. Masalan, mustasnoqlik ma'nosini bildiruvchi boshqa, bo'lak, o'zga, tashqari sinonimik qatorida boshqa, bo'lak, o'zga so'zlari predmet va shaxslarga, shuningdek, ish-harakatga nisbatan ham ishlatiladi; tashqari so'zi esa faqat narsa-predmet, shaxslarga nisbatangina qo'llanadi. Bunday holda so'z grammatik ma'nosini lingvistik bazada berish uchun grammatik ma'no ikki xil beriladi: 1) narsa va shaxslarga nisbatan mustasnoqlik, 2) ish-harakatga nisbatan mustasnoqlik. Shu asosda 2 xil grammatik ma'no shakllantirishga to'g'ri keladi.

birliklarini modellashirish va lingvistik bazalarni yaratish bo'yicha tadqiqotlar olib borilmoqda.

Maqolada o'zbek tilidagi yordamchi so'zlarning sinonimiyasi lingvistik bazaga kiritish jarayonidagi qiyinchiliklar tahlil qilinadi. Yordamchi so'zlarning grammatik ma'nosi va sintaktik vazifasiga qarab farqlanishi, ularning joylashuvi, grammatik qo'shimchalar talab qilishi va kontekstga bog'liq holda sinonimlik xususiyati o'zgarishi ilmiy nuqtayi nazardan o'rganildi. Ushbu tadqiqot natijalari mashina tarjimasi va tabiiy tilni qayta ishlash sohalariga uchun muhim ahamiyat kasb etib, tarjima sifatini oshirishga xizmat qiladi.

Natijalar shuni ko'rsatadiki, lingvistik bazani yaratishda yordamchi so'zlarning sintaktik va semantik xususiyatlari batafsil tahlil qilinishi, ularning leksik va grammatik omillar bilan bog'liqligi hisobga olinishi lozim. Bu esa, o'z navbatida, tabiiy tilni qayta ishlash tizimlarining yanada takomillashtirilishiga zamin yaratadi.

ADABIYOTLAR

- Elov, B. Til korpus matnlarini modellashirish va yashirin Dirix usuli. Toshkent: O'zMU nashriyoti, 2020, 4-son 45-47-betlar.
- Abdurahmonova, N. Морфологик таҳлилнинг машина таржимасида берилиши (от сўз туркуми модели мисолида). // Ўзбекистонда хорижий тиллар илмий методик электрон журнал №2-2017. Toshkent, 2017 p. 40-51
- Raab, M. Language model techniques in machine translation, Isroil, 2006, p.76-79
- Lembersky, G., Ordan N. Wintner S. Improving Statistical Machine Translation by Adapting Translation Models to Translationese //** University of Haifa, Israel, 2013
- Tantuğ, C., Adalı, E., Oflazer K., Lexical Ambiguity Resolution for Turkish in Direct Transfer Machine Translation Models, 2006, pp. 230-238.
- Sagay Z. A computer translation from English to Turkish. Journal of Computer linguistics, 1981, p. 426
- Hamzaoglu, I., Machine translation from Turkish to other Turkic languages and an implementation for the Azeri languages, Ankara, 1993, p.125
- Altıntaş, K., Turkish to Crimean Tatar Machine Translation System, Ankara, 2000, pp 129-131.
- Orhun M., Machine translation between uyghur language and turkish, Izmir, 2010, p.79
- Bozorova, S. Erkin kontekst grammatikasiga asoslangan sintaktik tahlil usullari. Filologiya fanlari doktori dissertatsiyasi. Toshkent, 2021, p. 88.
- Suyunova, M. O'zbek tilidagi grammatik omonimlarni modellashirish. O'zbek tilshunosligi tadqiqotlari, 3(1), 22-35.
- Xudayarova, S. So'z birikmalarini modellashirishning nazariy asoslari. O'zbek tili va adabiyoti, 2, 67-78.
- Oh, F. J., & Ney, H. Log-linear Modeling for Context-aware Machine Translation. Aachen University of Technology, Germany, 2014, p. 88-98.