

LEKSIK-SEMANTIK RAZMETKALASH NAZARIYASIDAGI QARASHLAR

Xayrulloeva Gulasal Fayzullo qizi

BuxDU o'zbek tilshunosligi va jurnalistika kafedrası o'qituvchisi

Abduqodirova Shoyista Tohir qizi

BuxDU talabasi

Annotatsiya: Mazkur maqolada korpus lingvistikasida semantik razmetka va semantik teglash muammolari va ularning yechimlari borasidagi fikr-mulohazalar bayon qilingan. Leksik ma'no (semema)ning tarkibiligi g'oyasi hozir ko'pchilik semasiologlar tomonidan tan olingan bo'lsa ham, lekin bu tarkibning xarakteri, ma'no komponentlarining bir-biriga munosabati masalasida hali ancha munozarali o'rinlar bor, shuningdek, maqolada semantik marker va semantik farqlovchi atamallari izohi ham keltirilgan.

Kalit so'zlar: korpus, semantik razmetkalash, semantika, semasiologiya, razryad, derevatsion tavsif, taksonomiya, mereologiya, topologiya, kauzatsiya, terminologik razmetka

Semantik razmetka (rus., razmetka) – til korpuslaridagi so'z yoki so'z birikmasining ma'lum bir semantik kategoriya yoki undan ham kichikroq semantik guruh (LSG, semantik maydon va to'da)ga mansubligini bildiruvchi, ma'noni maxsuslashtiruvchi belgi, izohlar majmui[1]. Korpusning semantik razmetkalari so'z ma'no(lar)i spetsifikasi, so'zning omonimlik, sinonimlik bilan bog'liq izohlar majmuini tuzish, so'zni kategoriyalash, uning mavzuviy guruh, LSG, semantik maydonga mansubligini belgilash, derivatsion xarakteristikasi, atash ma'nosi kabi belgilarni qamrab oladi. Rus tili milliy korpusi (NKRYA)ning semantik teg(razmetka)lash prinsiplarini kuzatish shuni ko'rsatdiki, bu korpusda bir so'zshaklga uch turdagi belgilar biriktiriladi[1]:

- 1) razryad (nom yoki unga ishora so'z: predmet, shaxs va h.);
- 2) avtomatik-semantik razmetkalash xarakteristika (leksema tegishli bo'lgan LSG yoki semantik maydon, kauzativlik belgilari va h.);
- 3) derivatsion tavsif.

Avtomatik-semantik razmetkalash quyidagi maydon bo'yicha guruhlanadi:

- 1) taksonomiya (leksema mansub LSG) – ot, sifat, fe'l, ravish so'z turkumlari uchun tegishli teg;

- 2) mereologiya ("butun-bo'lak"ka ishora qiluvchi belgi, "element-to'da/guruh"ga mansublik belgisi) – predmet va nopredmet atov birliklariga tegishli teg;

- 3) topologiya (ifodalanayotgan ob'ektning topologik mavqei) – narsa otlariga tegishli teg;

4) kauzatsiya – fe'llarga tegishli teg;

5) baho – predmet va nopredmet atov birliklari, sifat va ravishga tegishli teg.

Semantik razmetkalashda, boshqa razmetkalarda bo'lganidek, yagona standart shakl bo'lmasa ham, harf, raqam yoki faqat raqamdan iborat kodlardan foydalaniladi. Birinchi harf yoki raqam umumiy semantik ma'noni, keyingi belgi esa so'z ma'nosini yanada maxsuslashtiruvchi kichik semantik guruhni ifodalaydi. Semantik razmetka nafaqat so'z, balki ko'plab birikmalarni ham semantik guruhlariga birlashtiradi, bunday paytda turli birikuvdagi bir ma'noni bildiruvchi birikmalar bitta belgi bilan kodlanadi. Idiomatik birlik (ibora) tarkibidagi so'zlar miqdorini bildiruvchi axborot ham razmetkadan joy oladi. Semantik razmetka korpusdagi so'z ma'nosining ixtisoslashuvi, omonimlik, sinonimlik, ma'noviy guruhga ajratish kabi muammolarni hal qiladi. V.P.Zaxarov, S.YU.Bogdanovalar ham rus tili milliy korpusini tuzishda semantik razmetkalashning o'z variantini taklif qiladi[5]; ularning fikricha, har bir so'zga uch xil – so'z razryadi, avtomatik-semantik razmetkalash tavsif, derivatsion izoh berish deb aytilgani ma'qul.

Avtomatik-semantik razmetkalash korpus vositasida lingvistik amallar bajarish imkonini kengaytiruvchi ish quroli: korpusda semantik annotatsiyalangan birlikning mavjud bo'lishi faqat alohida so'zni emas, balki foydalanuvchi tomonidan so'ralgan birikma holdagi konstruktsiya qidiruviga javob berish imkonini paydo qiladi.

G.I.Kustova Rus tili milliy korpusi semantik razmetkasini tahlil qilar ekan, asosiy e'tiborni semantik razmetkaga asos bo'lgan "Leksikograf" bazasiga qaratadi[3]. Ushbu tizim Rossiya Fanlar akademiyasi Rus tili ilmiy tadqiqot institutida 1992-yilda ishlangan. Xususan, bazaning eng asosiy qismi sanalgan fe'llarning lingvistik ta'minoti E.V.Paducheva[4], narsa otlari E.V.Raxilina[5] tomonidan ishlab chiqilgan. Ushbu tizim oldindan semantik izohlar bilan shakllantirilgan so'zlarni ro'yxatga olishga mo'ljallangan lug'atdan iborat yuqori texnologiyali lingvistik baza hamda Korpusning avtomatik-semantik razmetkalash lug'ati sanaladi.

Semantik razmetka: belgi va guruhlar. Yuqorida aytib o'tilganidek, avtomatik-semantik razmetkalash morfologik razmetkaning davomi va kengaytmasi sanaladi; u o'zida uch guruh belgilar tizimini aks ettiradi[3]:

- 1) so'z yasovchi ko'rsatkichlar bilan shakllangan birlik;
- 2) avtomatik-semantik razmetkalash xususiyatga tegishli belgi: atoqli ot, narsa oti, asliy va nisbiy sifat; o'z va ko'chma ma'no;
- 3) xususiy-semantik belgi: mavzuviy (taksonomik) guruh, semantik maydon.

Zamonaviy avtomatik-semantik razmetkalashtizimlari lug'atlardan farqli o'laroq, semantik tarmoq yoki leksik tasniflarga bog'lanadi. Ular orasida Oxford Advanced Learners Dictionary (OALD)ga asosan ma'noni parchalab

qo'llovchi WordNet[4] mashhur. Ushbu avtomatik-semantik razmetkalashdasturi asosida ilk marta Braun korpusi razmetkalangan; dastur 234136 so'zshakl razmetkasini o'z ichiga olgan. Ushbu so'z qo'llash holatlarining 186575 tasi ko'p ma'noli. Keyinchalik LEXAS [Ng & Lee 1996] sistemasi ishlab chiqildi; unga qo'l razmetkasi tarzida ot va fe'lga tegishli nisbatan chastotasi yuqori 200 ta so'zning 192800 ifodasi kiritildi. Prinston universitetida yaratilgan SemCor korpusi 700000 so'zdan iborat, ularning 200000 tasi semantik jihatdan qo'l razmetkasi asosida WordNet 1.6 da razmetkalangan, WordNet 1.7.-2.0 da avtomatik qayta kodlangan.

Rus tili milliy korpusi (NKRYA) razmetkasi avtomatik-semantik razmetkalash tasnifi o'z maqsadi (lingvistlarni tadqiqotlar bilan ta'minlash, til haqidagi dalillarni ajratib olish) hamda kelib chiqishi (VINITI RANning fe'llarni turli ma'noda qo'llash modeli haqidagi talqin, ma'lumotni formatlashdan iborat "Leksikograf" leksik ma'lumotlar bazasining merosxo'ri hisoblanadi) bilan FrameNet tizimiga yaqin turadi. Ammo bu korpusning tasnifi FrameNetniki singari mukammal emas. Buning bir necha sababi mavjud[6]:

1) sinfosti guruhlardan iborat shajaraviy qidiruvdan farqli "to'g'ridan-to'g'ri" qidiruv tezkor natija beradi;

2) semantik sinf nomlari kompyuter oynasida bir vaqtning o'zida namoyon bo'lsa, lingvist-foydalanuvchi tasnif tizimidagi to'g'ri yo'nalishni tezroq topadi, shunga ko'ra, qidiruv savolini ham tezroq kiritadi. Semantik ko'p ma'nolilik masalasining yechimi ham leksik sinflar ko'lamida osonlashadi.

Korpus lingvistikasida semantik razmetkalash rivojlangan hamda mukammal razmetkalash dasturlari mavjud bo'lishiga qaramay, semantik razmetkalash jarayoni qator kamchiliklarga ega ekanligi ko'rinadi. M.YU.Zagorulko, I.S.Kononenko, E.A.Sidorova semantik razmetkaning xml, rtf, wiki formatda saqlanuvchi matnlarda mavjud bo'lishida quyidagi kamchiliklarni sanab beradi[2]:

1) istalgan fragment bo'lagini ajratib olish qiyin;

2) fragmentlar orasidagi aloqani ko'rsatish engil kechmaydi;

3) ko'p atributli ob'ektlarning tavsifini ajratib olish murakkab;

4) razmetkava atribut belgilari sistemasi teglovchi mutaxassis tomonidan mustaqil tanlanadi; semantik razmetkalar tizimining umumqabul qilingan standarti mavjud emas;

5) past samarali dasturiy ta'minot bilan ishlaydi.

Korpusni semantik razmetkalash dasturchi va tilshunosning kompetensiyalari birlashishi asosida ikki asosiy ko'rinishga ajratiladi:

1) terminologik razmetka – matndagi tushunchalarni nomlash bosqichi;

2) birliklararo munosabatni ko'rsatuvchi razmetkalar bilan ta'minlash.

Bunda har qaysi birlikning semantik maydondagi o'rni, boshqa birliklar bilan munosabatini aniqlash nazarda tutiladi.

O'zbek tili korpuslari avtomatik-semantik razmetkalashtamoyillari haqidagi birorta ishga duch kelmadik. Ammo B.Mengliyev va Sh.Xamroyevaning ayrim tadqiqotlarida semantik maydon, to'da va guruh tasniflarining avtomatik-semantik razmetkalashtizimini yaratishdagi o'zni haqida ayrim qaydlar mavjud[7].

Xulosa sifatida shuni ta'kidlash joizki, avtomatik-semantik razmetkalash tizimi umumlingvistik razmetka tizimi xususiyatlari bilan birga bir qancha xususiy belgilarga ham ega. Boshqa lingvistik razmetka tizimlariga o'xshash tomoni – razmetkaning lingvistik ta'minot yoki bazaga asoslanishi. Semantik razmetkaning o'ziga xos tomoni annotatsiyalashning ikki xil: faset va shajara usulining mavjudligida. Amalda birlikning xususiyatiga qarab bu ikki usuldan biri yoki ikkalasi aralash holda qo'llanadi. Faset usuli birlikka ketma-ket izoh berishni taqozo qilsa, shajara usuli birlik mansub bo'lgan guruh, to'da, maydonni belgilashni talab qiladi. Birinchi usulda so'zdan ma'noga, ikkinchi usulda ma'nodan so'zga qarab boriladi. Bu mutaxassisning razmetka jarayonida ham, interfeysda aks etadigan qidiruvda ham so'z → ma'no, ma'no → so'z prinsipining amal qilishini ko'rsatadi. Mavjud online korpuslarni kuzatish natijasida razmetka tizimida leksik-grammatik (1), leksik-tematik (2), uslubiy baho xarakteristikasi (3), derivatsion xarakteristika (4) kabi belgilarning ustuvorligiga guvoh bo'ldik. Konstanta va operator-klassifikator razmetkalar ham razmetka tizimidan doimiy joy oladigan razmetkahisoblanadi.

FOYDALANILGAN ADABIYOTLAR RO'YXATI:

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с. – С. 48-49.
2. Загорулько М.Ю., Кононенко И.С., Сидорова Е.А. Система семантической разметки корпуса текстов в ограниченной предметной области // <http://www.dialog-21.ru/media/1372/94.pdf>
3. Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. Москва: Индрик, 2005.- С. 155-174.
4. Кустова Г.И., Падучева Е.В. Словарь как лексическая база данных // ВЯ, 1994. – № 4.
5. Красылщик И.С., Рахилина Е.В. Предметные имена в системе «Лексикограф» // НТИ, сер. 2. – 1992. –№ 9. – С. 24-31.
6. Рахилина Е.В., Кобрицов Б.П., Кустова Г.И., Ляшевская О.Н., Шеманаева О.Ю. Лексико-семантическая разметка в национальном

корпусе русского языка. (Электрон ресурс):
<http://ruscorpora.ru/sbornik2005/10kustova.pdf>.

7. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Фил. фан. бўйича фалсафа док. (PhD) диссертацияси. – Бухоро, 2018. – Б. 73.