

WORDNETDA SEMANTIK MUNOSABATLARNING AKS ETISHI

Aynura Axmedova

ToshDO‘TAU Kompyuter lingvistikasi mutaxassisligi 1-kurs magistranti

E-mail: aynuraaxmedova345@gmail.com

Annotatsiya: Ushbu maqolada WordNetning til semantikasi, kompyuter lingvistikasi va tabiiy tilni qayta ishlash jarayonlaridagi o‘rni keng yoritilgan. WordNetning sinsetlar asosidagi konseptual tuzilishi, ular o‘rtasidagi asosiy semantik munosabatlar – sinonimiya, antonimiya, giponimiya, giperonimiya, meronimiya, xolonimiya tizim doirasida qanday formal modellashirilishi tahlil qilingan. Maqolada giponimiya va meronimiya munosabatlari WordNet brauzeridan real misollar bilan izohlangan.

Kalit so‘zlar: WordNet, semantik munosabatlar, sinset, giponimiya, giperonimiya, meronimiya, xolonimiya, antonimiya, leksik-semantik tarmoq, semantik iyerarxiya, ontologiya, semantik tahlil.

KIRISH

Zamonaviy lingvistika, kognitiv fanlar va kompyuter texnologiyalarining jadal rivojlanishi tilning mazmuniy tuzilishini tizimli ravishda modellashirishga bo‘lgan ehtiyojni keskin oshirdi. Ayniqsa, tabiiy tilni qayta ishlash (NLP), sun‘iy intellekt, ma‘lumotlarni semantik izlash, mashina tarjimai, bilimlarni tarmoqli shaklda ifodalash kabi murakkab jarayonlar leksik-semantik bilimlarni kompyuterga mos formal struktura ko‘rinishida taqdim etishni talab qiladi. Ana shunday ehtiyojlar tufayli WordNet kabi leksik-semantik tarmoqlar zamonaviy til texnologiyalari uchun asosiy lingvistik resurslarga aylandi.

WordNetning yaratilishi tilning ichki konseptual tuzilishini kompyuter modeli sifatida aks ettirishga qaratilgan ilk fundamental loyihalardan biri bo‘lib, u leksik birliklarning semantik munosabatlarga asoslangan tarmoqli tuzilmasini taklif etdi. Sinsetlar, giponimiya, meronimiya, antonimiya, xolonimiya kabi semantik munosabatlar orqali WordNet insonning mental leksikonidagi ma‘no munosabatlarini aniq grafik struktura sifatida modellashiradi. Bu esa tilning semantik qatlamini mashina tomonidan qayta ishlanishi mumkin bo‘lgan shaklga keltirish imkonini beradi.

ASOSIY QISM

WordNet bu ingliz tilidagi leksik birliklarning ma‘nolarini formal semantik tarmoqlar shaklida tashkil qiluvchi elektron leksik ma‘lumotlar bazasidir. Uning asosiy maqsadi so‘zlarning tashqi shaklidan ko‘ra, ularning konseptual mazmuni va semantik munosabatlarini aks ettirishdan iborat. Shu bois WordNet an‘anaviy lug‘atlardan farqli o‘laroq, so‘zlarni alifbo tartibida emas, balki *sinset* deb ataladigan ma‘no birliklari asosida guruhlaydi. Sinset bir xil ma‘noni ifodalovchi sinonimlar to‘plami bo‘lib, WordNetning eng kichik ma‘nodagi birligi hisoblanadi. Har bir sinset ma‘lum bir konseptni aks ettiradi va u o‘ziga xos semantik munosabatlar orqali boshqa sinsetlar bilan bog‘lanadi [2]. Aynan shu ma‘noviy bog‘lanishlar – giponimiya, meronimiya, antonimiya, giperonimiya va boshqalar WordNetning ishlash mexanizmini tashkil qiladi hamda tilning mazmuniy strukturasini tarmoqli grafik ko‘rinishida modellashirishga imkon beradi.

WordNetning ishlash prinsipi quyidagi g‘oyaga asoslanadi:

tilning leksik-semantik tizimi inson mental leksikoniga o‘xshash tarzda konseptlar va ular o‘rtasidagi munosabatlar orqali iyerarxik va tarmoqli shaklda tashkil topgan. WordNet aynan shu kognitiv modelni kompyuterlashtirilgan ko‘rinishda taklif etadi. Shuning uchun u nafaqat lug‘at, balki semantik tarmoq, bilim bazasi va psixolingvistik model vazifasini ham bajaradi.

WordNetdagi har bir so‘z bir yoki bir nechta sinsetlarga mansub bo‘lishi mumkin. Bu hol ko‘p ma‘nolilikni formal tarzda ifodalash va har bir ma‘noni alohida semantik tugun sifatida qayd etish

imkonini beradi. Sinsetlar orasidagi aniq belgilangan munosabatlar esa WordNetni NLPdagi ko‘plab jarayonlar: so‘z ma’nosini aniqlash, semantik o‘xshashlikni hisoblash, avtomatik tarjima, semantik qidiruv kabi vazifalar uchun samarali lingvistik resursga aylantiradi.

WordNet tizimida ingliz tilidagi leksik birliklar to‘rt asosiy so‘z turkumiga ajratilgan holda beriladi: *otlar* (nouns), *fe’llar* (verbs), *sifatlar* (adjectives) va *ravishlar* (adverbs). Tizimga faqat mazmun yuklaydigan so‘zlar kiritilgan bo‘lib, artikl, ko‘makchi, bog‘lovchi kabi so‘z turkumlari WordNet tarkibiga qo‘shilmagan. Buning sababi shundaki, WordNetning asosiy maqsadi ma’noli birliklar orasidagi semantik munosabatlarni ko‘rsatishdir. Shu bois tizim mazmun markazida turuvchi to‘rt so‘z turkumini qamrab oladi va ularning har biri alohida semantik tarmoq sifatida tashkil etadi.

WordNetdagi eng asosiy semantik munosabatlardan biri **sinonimiya** bo‘lib, u bir ma’noni ifodalovchi so‘zlarni *sinset* deb ataladigan maxsus birlikka birlashtirish orqali beriladi. Har bir sinset bitta tushunchani ifodalaydi va uning tarkibida o‘zaro sinonim bo‘lgan ikki yoki undan ortiq so‘z joylashadi. Masalan, *katta*, *ulkan*, *gigant* kabilar WordNetda bitta sinset tarkibida qayd etiladi. Quyida Princeton Wordnetda *big* so‘zining sinonimlari ro‘yxati berilgan. Avvalo, Wordnetda **big** so‘zi kiritilganda uning sifat so‘z turkumidagi 13 ta ma’nosi (semasi), *ravish* so‘z turkumidagi 4 ta ma’nosi (semasi) va har bitta ma’nosining sinonimlarini ham berildi. Big (katta) so‘zining o‘lcham, son, miqdor, hajm yoki ko‘lam jihatidan odatdagidan ortiq, yuqori darajada bo‘lgan narsa: “*katta shahar*”, “*katta xona*” ma’no (sema)siga berilgan sinonimlar ro‘yxati:

- => *ample, sizable, sizeable* (anchagina katta; yetarlicha katta);
- => *astronomic, astronomical, galactic* (behad katta; tasavvur qilgudek bo‘lmagan darajada ulkan);
- => *bear-sized* (ayiqdek katta; ayiq o‘lchamidagi);
- => *bigger, larger* (narsaga nisbatan kattaroq);
- => *biggish, largish* (bir oz katta, yirikroq);
- => *blown-up, enlarged* (kattalashtirilgan; kengaytirilgan);
- => *bouffant, puffy* (shishiq, qavariq; soch turmagi yoki kiyim haqida);
- => *broad, spacious, wide* (keng; juda katta maydonni egallagan);
- => *bulky* (hajmi kattayu, og‘ir bo‘lmagan; kurtakdek katta);
- => *capacious* (keng sig‘imli; katta hajmga ega);
- => *colossal, prodigious, stupendous* (juda-juda katta; hayratlanarli darajada ulkan);
- => *deep* (miqdor yoki hajm jihatidan katta);
- => *double* (ikki kishilik o‘lchamdagi);
- => *enormous, tremendous* (favqulodda katta; nihoyatda ulkan);
- => *cosmic* (kosmos miqyosidagi; tasavvurdan tashqari katta);
- => *elephantine, gargantuan, giant, jumbo* (fildek ulkan; juda massaviy);
- => *epic, heroic, larger-than-life* (epik darajada ulkan; juda hayratlanarli o‘lchamda);
- => *extensive, extended* (keng ko‘lamli; katta hududni qamrab oluvchi);
- => *gigantic, mammoth* (gigant darajada katta; mamontdek ulkan);
- => *great* (turiga nisbatan katta; yirik);
- => *grand* (ko‘lami katta; ulkan ko‘rinishli);
- => *huge, immense, vast* (favqulodda katta; g‘oyat ulkan);
- => *hulking, hulky* (yirik, qo‘pol kattalikdagi);
- => *humongous, banging, thumping, whopping, walloping* (juda katta; norasmiy tarzda “ultrabig” ma’nosida);
- => *king-size, king-sized* (eng katta standart o‘lcham; “king-size”);
- => *large-mouthed* (katta og‘izli (biologik tavsif);
- => *large-scale* (keng ko‘lamli: hujum, loyiha);
- => *large-scale* (keng masshtabda chizilgan: xarita);
- => *life-size, lifesize, life-sized, full-size* (asl nusxasiga teng o‘lchamda; real kattalikda);
- => *macro* (katta miqyosli; makro darajada);
- => *man-sized* (erkak odam o‘lchamiga mos katta);
- => *massive, monolithic, monumental* (yirik massaviy, og‘ir; monolit darajada katta);

- => *massive* (yirik ko‘lamli; keng miqyosli o‘zgarishlar bilan bog‘liq);
- => *medium-large* (o‘rtacha kattadan biroz katta);
- => *monstrous* (oddagidan keskin katta; g‘ayritabiiy kattalik);
- => *mountainous* (tog‘dek ulkan; tog‘ sifat katta);
- => *outsize, outsized, oversize, oversized* (standartdan katta; oshirilgan o‘lchamdagi);
- => *overlarge, too large* (ortiqcha katta; haddan ziyod, yirik);
- => *plumping* (o‘z turiga nisbatan juda katta);
- => *queen-size, queen-sized* (king-sizedan kichikroq, lekin yirik standart o‘lcham);
- => *rangy* (harakatlanish uchun keng joy taqdim etuvchi; katta maydonli);
- => *super* (nihoyatda katta; “super yirik”);
- => *titanic* (ulkan kuch yoki hajmga ega; titan kabi);
- => *volumed* (shaklan kattalashtirilgan, ko‘p bo‘lakli, qavariq shaklli);
- => *voluminous* (katta hajmli; keng, ko‘p matoli);
- => *whacking* (britancha: nihoyatda katta; juda yirik);
- => *wide-ranging* (ko‘lami keng; ko‘p narsani qamrab oluvchi).

Big (katta) so‘zining **antonimi** (qarama-qarshi ma‘no) sifatida *small, little* (kichik, mayda) so‘zlari berilgan.

WordNetdagi eng muhim semantik munosabatlardan biri – **giponimiya** bo‘lib, u “is-a” tamoyiliga asoslangan umumiylik-xususiylik munosabatini ifodalaydi. Bu leksemalar o‘rtasidagi jins-tur munosabatidir. Bu keng (umumiy) tushunchani ifodalovchi va tor (xususiy) tushunchani ifodalovchi leksemalararo munosabat turidir. Keng, umumiy, jins tushunchasini ifodalovchi leksema ma‘nosi tor, xususiy, tur tushunchasini ifodalovchi leksema ma‘nosini qamrab oladi [1]. Masalan, “meva” va “olma” leksemalari o‘rtasida jins-tur munosabati mavjud. “Meva” keng, umumiy jins tushunchasini bildiradi, “olma” esa shu toifaning tor, xususiy turi hisoblanadi. Ya‘ni “meva” ma‘nosi “olma” tushunchasini o‘z ichiga oladi. Wordnetda *olma* so‘zining **giperonimlari** quyidagicha:

- => *apple* (olma);
- => *edible fruit* (yeyiladigan meva);
- => *fruit* (meva);
- => *reproductive structure* (ko‘payish tuzilmasi);
- => *plant organ* (o‘simlik organi);
- => *plant part / plant structure* (o‘simlik qismi / o‘simlik tuzilmasi);
- => *natural object* (tabiiy obyekt);
- => *whole / unit* (butunlik / birlik);
- => *object / physical object* (obyekt / jismoniy obyekt);
- => *physical entity* (fizik mavjudot);
- => *entity* (mavjudot);

WordNet brauzeri giponimlarni ikkita formatda ko‘rsatadi: **qisqa** (brief) va **to‘liq** (full). Qisqa ko‘rinish faqat birinchi darajali giponimlarni –so‘zning bevosita ostida joylashgan turlarini ko‘rsatadi. Bunda semantik iyerarxiyaning pastki qatlamlari ochilmaydi. Qisqa ro‘yxat faqat asosiy turlarni tez ko‘rish uchun mo‘ljallangan.

Masalan, *apple* (olma) so‘zi uchun **qisqa giponimlar**:

- => *crab apple, crabapple* – mayda, nordon olma; asosan murabbo yoki jele tayyorlashga mos;
- => *eating apple, dessert apple* – asosan xom holda iste‘mol qilish uchun yetishtiriladigan shirin olma;
- => *cooking apple* – asosan pishirish (pishiriqlar, piroglar, olmadan tayyorlanadigan souslar va boshqalar) uchun ishlatiladigan olma turi.

apple (olma) so‘zi uchun **to‘liq giponimlar**:

- => *crab apple, crabapple* –(mayda, juda nordon olma; murabbo va jele tayyorlashga mos);
- => *eating apple, dessert apple* –(asosan xom yeyiladigan shirin olma);
- => *Baldwin* – (qizil yoki sariq-qizil po‘stli amerikalik olma navi);

- => *Cortland* –(katta, qizil po‘stli olma turi);
- => *Delicious* –(shirin ta‘mli mashhur olma navi);
 - => *Golden Delicious, Yellow Delicious* –(sarg‘ish po‘stli shirin nav);
 - => *Red Delicious* –(yorqin qizil po‘stli, dunyoda ko‘p yetishtiriladigan navi);
- => *Empire* –(MacIntoshga o‘xshash; xom va pishirish uchun mos);
- => *Grimes' golden* –(kech kuzda pishadigan sariq olma);
- => *Jonathan* –(kech pishadigan qizil olma; xom iste‘mol qilinadi);
- => *McIntosh* –(erta pishadigan mashhur olma navi);
 - => *Macoun* –(McIntoshga o‘xshash, juda sharbati ko‘p);
- => *Northern Spy* –(katta, kech pishadigan, qizil-sariq chiziqli po‘stli olma);
- => *Pearmain* –(qizil po‘stli bir nechta navlar guruhi);
- => *Pippin* –(sariq yoki yashil-sariq, usti qizil bo‘lgan yuqori sifatli navlar guruhi);
 - => *Cox's Orange Pippin* –(xushbo‘y, mashhur Pippin navi);
- => *Prima* –(asosan xom yeyiladigan olma turi);
- => *Stayman* –(Shenandoah vodiysida yetishtiriladi);
- => *Winesap* –(qoramtir qizil po‘stli, qattiq va xushbo‘y olma);
- => *Stayman Winesap* –(qattiq, nordon-shirin; xom va pishirish uchun mos);
- => *Granny Smith* –(yashil po‘stli, qattiq va nordon olma turi);
 - => *cooking apple* –(asosan pirog tayyorlash uchun ishlatiladigan olma);
- => *Bramley's Seedling* –(juda yirik pishirish olmasi);
- => *Lane's Prince Albert* –(asosan pishirish uchun mo‘ljallangan navi);
- => *Newtown Wonder* –(pishirish uchun juda mos nav);
- => *Rome Beauty* –(katta, qizil; pishiriqlar uchun ishlatiladi);

Xolonimiya (partonomiya) – “x y ning qismi” munosabati bo‘lib, bunda x butun, y esa uning qismi sifatida qaraladi. Partonomiya bu til birliklari o‘rtasidagi “butun va qism” munosabatidir. Butun o‘z qismlarining birligidan tashkil topadi. Qism esa butunning tarkibiy bo‘lagi, uni tashkil etuvchi element hisoblanadi. Obyektiv olam yaxlit butunlik (sistema) tabiatiga ega bo‘lib, u bosqichma–bosqich qismlarga ajraladi. Butun olam galaktikadan tortib elementar zarrachalargacha “butun–qism” munosabatida mavjud bo‘ladi. Olamning akslanishi bo‘lgan tilda ham butun–qism munosabatlari muhim ahamiyat kasb etadi va tilning qismlardan iborat pog‘onali tuzilishga ega yaxlit tizim ekanligini ko‘rsatadi: matn – gap, gap – so‘z birikmasi, so‘z birikmasi – so‘z, so‘z – morfema, morfema – fonema [1].

WordNet **xolonimlarni** ikki xil formatda ko‘rsatadi:

1. *Regular (asosiy) xolonimlar*. Bu ko‘rinish faqat to‘g‘ridan-to‘g‘ri butun–qism munosabatlarini chiqaradi. Ya‘ni leksema bevosita qaysi yaxlit narsa tarkibiga kirishi ko‘rsatiladi. Masalan: *apple* → *apple tree* (olma → olma daraxtining qismi). Bu faqat birinchi pog‘ona holonimlarni beradi.

2. *Inherited (kengaytirilgan) xolonimlar*. Bunda leksema qanday kengroq, yuqori darajadagi butunlar tarkibiga kirishi ham ko‘rsatiladi. Bu shakl iyerarxik daraxt bo‘ylab yuqoriga ko‘tarilgan butunlarni ham chiqaradi. Masalan:

apple → *apple tree* → *orchard* → *ecosystem* → *biosphere*
(olma → olma daraxti → bog‘ → ekotizim → biosfera)

WordNet **meronimlarni** ikki shaklda beradi:

1. *Regular (asosiy) meronimlar*. Bu ko‘rinish leksemaning faqat bevosita tarkibiy qismlarini ko‘rsatadi. Masalan: *apple* → *core* (olma → mag‘iz), *apple* → *flesh* (olma → go‘sht qismi), *apple* → *peel* (olma → po‘stloq).

2. *Inherited (kengaytirilgan) meronimlar*. Bu shaklda esa nafaqat olmaning asosiy qismlari, balki ularning ham tarkibiy bo‘laklari (sub-meronimlar) ko‘rsatiladi. Ya‘ni semantik iyerarxiya chuqurroq ochiladi. Masalan: *apple* → *peel* (olma → po‘stloq), *peel* → *plant tissue* (po‘stloq → o‘simlik to‘qimasi), *plant tissue* → *plant cell* (o‘simlik to‘qimasi → o‘simlik hujayrasi), *plant cell* → *organelle* (o‘simlik hujayrasi → organoid), *organelle* → *molecule* (organoid → molekula).

XULOSA

WordNet zamonaviy leksik-semantik modellashning eng muhim va muvaffaqiyatli loyihalaridan biri bo'lib, tilning konseptual, tarmoqli va iyerarxik tuzilishini formal tarzda qayta yaratishga imkon beradi. Synsetlarga asoslangan bu model leksik birliklarni faqat alifbo tartibida emas, balki mazmuniy munosabatlar orqali birlashtirib, inson mental leksikonidagi semantik bog'lanishlarning shaklini aks ettiradi. WordNetdagi asosiy semantic munosabatlar – sinonimiya, antonimiya, giponimiya, meronimiya, xolonimiya va boshqa munosabatlar til birliklari o'rtasidagi konseptual aloqalarni aniqlash, tasniflash va kompyuterda qayta ishlash uchun zarur bo'lgan aniq strukturani yaratadi. WordNetning giponim-giperonim zanjirlari tushunchalarning umumiylik-xususiylik munosabatlarini formal shaklda namoyon etsa, meronim-xolonim munosabatlari butun-qism aloqalarini iyerarxik ravishda ifodalaydi. Bu ikki asosiy munosabat turi WordNet semantik grafigning tayanch qatlamini tashkil qiladi. Sinsetlar o'rtasidagi mazkur bog'lanishlar semantik o'xshashlikni aniqlash, avtomatik tarjima, matnni semantik izlash, so'z ma'nosini aniqlash kabi NLP vazifalarida katta amaliy ahamiyatga ega hisoblanadi.

FOYDALANILGAN ADABIYOTLAR:

1. Hakimova M. Semasiologiya. –Toshkent, (2008); –B. 79
2. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. Introduction to WordNet: An on-line lexical database. International journal of lexicography, 1990. –p. 3
3. Miller G.A. WordNet: An on-line lexical database. International Journal of Lexicography, 1990;
4. Fellbaum C. WordNet: An Electronic Lexical Database. MIT Press, 1998;
5. Fellbaum C. WordNet. In: Theory and Applications of Ontology, 2010;
6. Resnik P. Using information content to evaluate semantic similarity. Proceedings of IJCAI, 1995;
7. Banerjee S., and Pedersen T. Extended gloss overlaps for sense disambiguation. Proceedings of CICLing, 2003;
8. Lesk, M. (1986). Automatic sense disambiguation using dictionary overlaps. SIGDOC Conference Proceedings.
9. Pedersen T. WordNet-based similarity measures. University of Minnesota Technical Reports, 2004;
10. Pianta E., Bentivogli A., and Girardi C. MultiWordNet: Developing an aligned multilingual WordNet. Proceedings of GWC, 2002;

Internet saytlar:

1. <https://uz.wikipedia.org>
2. <https://wordnet.princeton.edu/>