

2-25-2020

SEMANTIC TAG CATEGORY IN CORPUS LINGUISTICS: EXPERIENCE AND ANALYSIS

Dildora Axmedova Bahodirovna
Basic doctoral student of Bukhara State University

Follow this and additional works at: <https://uzjournals.edu.uz/buxdu>



Part of the [Linguistics Commons](#)

Recommended Citation

Bahodirovna, Dildora Axmedova (2020) "SEMANTIC TAG CATEGORY IN CORPUS LINGUISTICS: EXPERIENCE AND ANALYSIS," *Scientific reports of Bukhara State University*: Vol. 4 : Iss. 1 , Article 15. Available at: <https://uzjournals.edu.uz/buxdu/vol4/iss1/15>

This Article is brought to you for free and open access by 2030 Uzbekistan Research Online. It has been accepted for inclusion in Scientific reports of Bukhara State University by an authorized editor of 2030 Uzbekistan Research Online. For more information, please contact sh.erkinov@edu.uz.

5. Ширинова М.Ш. Ортология адабий меъёр ҳақидаги фан сифатида. //БухДУ илмий ахбороти, 2019 йил 3-сон. 127-131-б.

6. Ширинова М.Ш. Кино тили тушунчаси ва унинг ўзига хос хусусиятлари. // БухДУ илмий ахбороти, 2019 йил, 4-сон.

7. Shirinova M.Sh. The concept of film language and its studiedness in linguistic aspect. Актуальные вопросы современной науки и образования: сборник статей Международной научно-практической конференции. В 2 ч. Ч. 1. – Пенза: МЦНС “Наука и Просвещение”, – 2020. – С. 157-159.

https://studref.com/473325/zhurnalistika/yazyk_kino_yazyk_ekrana

УДК: 410

КОРПУС ЛИНГВИСТИКАСИДА СЕМАНТИК ТЕГ КАТЕГОРИЯСИ: ТАЖРИБА ВА ТАҲЛИЛ

КАТЕГОРИЯ СЕМАНТИЧЕСКИЙ ТЭГ В КОРПУСНОЙ ЛИНГВИСТИКЕ: ЭКСПЕРИМЕНТ И АНАЛИЗ

SEMANTIC TAG CATEGORY IN CORPUS LINGUISTICS: EXPERIENCE AND ANALYSIS

Ахмедова Дилдора Баҳодировна
БухДУ таянч докторанти

Аннотация. Мақолада корпус бирликларини семантик аннотациялашнинг муҳим категориялари саналган семантик разметка, семантик разметка тизими, оператор, классификатор, кўпмаънолилик, семантик филтрнинг ўзига хос хусусияти ҳамда имкониятлари ҳақида баҳс юритилади. Семантик тег тизимига турли ёндашув ва қарашлар ҳақида фикр билдирилган, семантик теглаш муаммолари, кўпмаънолилик ва омонимлик муаммосини ечиш ҳамда филтрлашга доир айрим мулоҳазалар; рус, калмиқ, бошқирд, чуваш тили корпусининг семантик разметкасига доир масалалар таҳлили келтирилган.

Таянч сўзлар: семантик разметка, семантик тег, разметка тизими, оператор, классификатор, кўпмаънолилик, семантик филтр.

Аннотация. В данной статье рассмотрены особенности и возможности важнейших категорий семантической аннотации корпусных единиц, таких как, семантическая разметка, система семантической разметки, оператор, классификатор, многозначность, семантический филтр. Высказано мнение о различных подходах и взглядах на систему семантических тегов, приведены рассуждения о пути решения проблемы семантических тегов, их многозначности и омонимии, а также представлены некоторые рассуждения, касающиеся семантических филтров. Проведён анализ задач семантической разметки языкового корпуса русского, калмыцкого, башкирского и чувашского языков.

Ключевые слова: семантическая разметка, семантический тег, система разметки, оператор, классификатор, многозначность, семантический филтр.

Annotation. This article discusses the semantic markup system, the semantics markup, the operator, the classifier, the versatility, the properties and capabilities of semantic filtering, which are important categories for the semantic annotation of corpus units. Different approaches and theories to semantic tag system are expressed; some considerations for solving and filtering semantic touch problems, versatilities and homonymic; an analysis of the semantic circuitry of the Russian, Kalmyk, Bashkir, Chuvash language is given.

Key words: semantic markup, semantic tag, markup system, operator, classifier, versatility, semantic filter.

Кириш. Разметканинг исталган тури лингвистик ва дастурий таъминотни талаб қилганидек, семантик разметка ҳам лексик/лексикографик/семантик базалар асосида амалга оширилади. Мақолада рус тили миллий корпуси семантик разметкасига асос бўлган

“Лексикограф” тизими имконияти ўрганилган, таҳлил қилинган. Шунингдек, корпус разметкаси учун лингвистик таъминот вазифасини бажарувчи семантик тасниф, семантик луғатлар ҳақида маълумот берилган.

Семантик аннотациялашнинг икки хил: фасет ва шажара усули фарқланади; ушбу икки усулнинг ўзига хос томонлари, турли бирликларни разметкада ҳар бир усулнинг аҳамияти очиқ берилган. Амалда бирликнинг хусусиятига қараб бу икки усулдан бири ёки иккаласи аралаш ҳолда қўлланади. Фасет усули бирликка кетма-кет изоҳ беришни тақозо этса, шажара усули бирлик мансуб бўлган гуруҳ, тўда, майдонни белгилашни талаб қилади. Биринчи усулда сўздан маънога, иккинчи усулда маънодан сўзга қараб борилади. Бу мутахассиснинг разметка жараёнида ҳам, интерфейсида акс этадиган қидирувда ҳам сўз —> маъно, маъно —> сўз принципining амал қилишини кўрсатади. Мавжуд online корпусларни кузатиш натижасида разметка тизимида лексик-грамматик, лексик-тематик, услубий баҳо характеристикаси, деривацион характеристика каби белгиларнинг устуворлиги; константа ва оператор-классификатор теглар ҳам разметка тизимидан доимий жой оладиган теглиги аниқланган. Ўзбек тили корпусларини семантик разметкада ушбу разметка тизимидан фойдаланишга доир таклифлар келтирилган.

Асосий қисм. Семантик тег тизими: турли ёндашув ва қарашлар. Корпус лингвистикасида семантик разметка, унинг теглар тизими, тег категориялари, семантик теглаш муаммолари, кўпмаънолилиқ ва омонимлик муаммосини ҳал этиш масаласига оид қатор тадқиқотлар вужудга келган. Жумладан, В.В. Куканова[5], А.А. Кретов[4] Г.И. Кустова, О.Н. Ляшевская, Е.В. Падучева, Е.В. Рахилина, Б.П. Кобрицов, ва Т.И. Резникова[6] шу масалалар ечимига бағишланган қатор ишларни эълон қилганлар. Қайд этиб ўтилган олимлар рус, қалмиқ, бошқирд тили корпусларининг семантик разметкасига доир фикр билдирганлар; корпус семантик параметр асосидаги қидируви, семантик разметкаси таҳлил қилинган. Қуйида уларнинг семантик теглар тизимида олимларнинг ёндашувига тўхталамиз.

В.В. Куканова “Принципы семантической разметки национального корпуса калмыцкого языка” номли мақоласида қалмиқ тили миллий корпуси дастурий таъминоти таркибига график ва морфологик анализатор, корпус менежери киришини, ушбу корпуснинг семантик разметкаси рус тили миллий корпуси[15] асосида ишлаб чиқилганлигини таъкидлайди[5]. Корпус семантик разметка тизимини ишлаб чиқишда бошқирд тили корпуси[1] тажрибаси, С.Г. Бархударов таҳрири остидаги рус тили семантик луғати[3], шундай турдаги Н.Ю. Шведова луғати[13]даги таксономик гуруҳлар таснифига таянилган. Қалмиқ тили миллий корпуси семантик разметкаси муаммосини ҳал этишда Рус тили миллий корпусининг семантик разметкаси маълумотнома вазифасини бажарган. Рус тили миллий корпуси семантик разметкаси эса “Лексикограф” системаси лингвистик базаси воситасида амалга оширилган[7].

В.В. Куканова “Семантик разметка нега керак?” деган саволга С. Гиндиннинг сўзлари билан жавоб беради: “Тил тавсифига икки томондан ёндашиш мумкин: шаклдан мазмунга ҳамда мазмундан шаклга. Биринчи ёндашув маълум тил бирлигининг мавжуд барча маъноларини имкон қадар аниқ ва тўлиқ тавсифлашни, унинг қўлланиш контекстини кўрсатади. Иккинчи ёндашувда эса, аксинча, айни бир мазмунни ифодалай оладиган барча шаклларни бир нуқтага йиғиш мақсад қилинади: бунда маълум маъно мажмуи тўпланади. Иккала ёндашув ҳам тилшунослик ва тил таълими учун бирдек аҳамиятли.

Бундан ташқари, грамматика ва луғат фойдаланувчи/ўрганувчиси ҳам она тили/хорижий тилни ўрганар экан, икки хил ёндашувга ҳам эҳтиёж сезади. Агар нотаниш сўз ёки конструкция учраса, луғат/грамматикага мурожаат қилинади; киши ўзи нимадир ҳақида ёзиши, гапириши лозим бўлса, унда сўз/ифода қидириб қолади: нима демоқчилигини билади, ҳис қилади, аммо керакли сўз/грамматик шаклни топа олмайди”[5]. Ҳақиқатан ҳам, семантик разметка тил лексикаси тадқиқи учун жуда муҳим: тадқиқотчи/фойдаланувчи/ўқувчига сўзнинг бирикиш имконияти, синтактик конструкциясини кузатишга кенг йўл очади. Масалан, айрим бир сифатлар борки, фақат маълум отлар билан семантик валентлик ҳосил

кила олади. Интерфейсда мана шундай сўровларни ҳосил қилиш мисоллар массиви орқали сўзнинг валентлигини кўрсатади.

Мутахассислар семантик аннотациялашнинг икки хил: фасет ва шажара усулини фарқлайдилар[5]. Одатда, корпус семантик разметкасида икки усулдан бирдек фойдаланишади. Сўз бирдан ортиқ маънони ифодалаши мумкин, шу сабабли доим ҳам шажара усулида таснифлаш тўғри натижа бермайди, кўп маъноли сўзларни тавсифлашда фасет усули қулай. Масалан, *шайтон* сўзи контекстга қараб қуйидаги маъноларини билдиради[14]:

ШАЙТОН I [^{1a} *شیطان* - иблис; жин, ажина] **1 дин. Одамларни дин йўлидан оздирувчи, уларни гуноҳга, жиноятга, разолатга бошловчи ёвуз рух ёки ёвуз рухларнинг бошлиғи; иблис.** *Эй улуг шоҳ, бу йўлдан қайт, шайтон васвасасига учма, кейин пушаймон бўлиб юрма. «Олтин бешик». Бойни ҳам шайтон йўлдан урган. К. Яшин, Ҳамза. Ибодат қилмаслик умуман гуноҳ эмас, лекин ибодат қилмаган киши, шайтоннинг гапига кириб қилмагани учун, гуноҳ ҳисобланади. А. Қаҳҳор, Қабрдан товуш. – Оббо, қизиталоқ. Шайтоннинг домласи экан-а! Зақунчиларнинг кўпи шунақа бўлади, – деди Тантубойвачча папиросни тортиб. Ойбек, Танланган асарлар.*

2 кўчма Айёр, қув одам; алдамчи. *Анави магазинчи бор-ку. Ўлгудек шайтон.. Р.Файзий, Чўлга баҳор келди.*

3 кўчма Шўх, ўйноқи. *Қиз бола шайтон бўлади-да, тавба, бўй-баста чўзилиб, ростдан ҳам, суқсурдек қиз бўлибди. Ш. Тошматов, Эрк куши. – Ҳа, шайтон қиз, – деди Ўринбой, – шу ерда ҳам қўймайсан-а? Ойдин, Келин ўғил туғибди. Бизнинг Авазга ўхшаган битта шайтонроғи «Йигирма олтинчиси отилмасин!» деб қийқиради. С. Сиёев, Ёруғлик. Ноумид – шайтон. Инсон доим умид, яхши ният билан яшашини, умидсизланмасликни қайд этувчи, шунга даъват этувчи ибора. Дийдор кўришамиз, ноумид – шайтон! И. Раҳим, Чин муҳаббат. – Кўйинг, ноумид – шайтон, – Ойшабону эрига таскин берган бўлди. Ҳ. Гулом, Машғал. Шайтон арава с.т. эск. айн. велосипед. Йиртиқ-ямоқ кийим билан ўсган ўғиллари шайтон аравада ўтиб, атроф маҳалла болаларининг кўзларини ўйнатди. Ойбек, Танланган асарлар. Шайтон йиғиси Ёлғондакам, йўлига бўладиган йиғи. Озгина «шайтон йиғиси» қилмагунингча, иш ўнганмайди. А. Қодирий, Улоқда. У буни «узатиладиган қизларнинг шайтон йиғиси» деб тушунди. Ойбек, Танланган асарлар. Шайтонга ҳай бермоқ Бирон хатти-ҳаракатдан ўзини тиймоқ, ўзини босмоқ[Ҳайитжон]. Айбмас, хўп андиша қилгин, шайтонга ҳай бергин, бу кунлар ҳам ўтар. Ҳамза, Паранжи сирлари. Бир тўполон чиқазиб, обрўсини уч пул қилай дедим-у, яна шайтонга ҳай бердим. «Муштум». Шайтони зўр (ёки кучли) Шўхлиги («шайтонлиги») тез ва кучли қўзғаладиган, шундай хусусиятли. Яна ким билсин, бу қизларни! Уларнинг шайтони зўр бўлади, дейишади. И. Раҳим, Чин муҳаббат. Шайтони қўзимок «Шайтони» қўзғамок, жини тутмоқ[Холида]. Қўл чўзса етгудай жойда бекиниб ётган акаси билан Мунавварни кўрмади. Мунавварнинг шайтони қўзиб: – Этагидан торт! – деди Йўлдошга. С. Анорбоев, Меҳр.*

Кўриниб турибдики, ўзбек тилининг изоҳли луғатида *шайтон* сўзининг учта маъноси берилган бўлиб, учинчи – кўчма маъносида бирикма (эркин ва турғун) бирликлар ҳам мавжуд. Дейлик, кўп маъноли сўзларнинг маъноларини фарқлаш филтер ёрдамида автоматик амалга оширилди, аммо бирикма бирликларни компьютер қандай “тушунсин” ва сўровга “жавоб берсин”? Бундай бирликларни семантик теглашнинг бир неча ечимини таклиф қилиш мумкин. Ибора, ўз маънодаги турғун бирикма, таркибли атоқли ва турдош отлар учун алоҳида лингвистик таъминот ишлаб чиқиш лозим. Бу жараённинг алгоритми тахминан қуйидагича бўлади:

1) юқорида ажратиб кўрсатилганидек, изоҳли луғатларда мавжуд маънолари йиғилади ҳамда лингвистик таъминот тузилади;

2) бирикма леммаланади, яъни битта лемма (сўз)га тенглаштирилади. Масалан: <ноумид – шайтон>, <шайтон арава>, <шайтон йиғиси>, <шайтонга ҳай бермоқ>, <шайтони

зўр>, <шайтони кўзимок>. Энди компьютер буни битта сўз деб тушунади ва шундай сўров бўлганда лингвистик таъминот базасидан ўз семантик изоҳи орқали жавоб топади.

Шайтон леммасининг семантик аннотацияси қуйидаги кўринишда бўлади:

<*Шайтон*> Одамларни дин йўлидан оздирувчи, уларни гуноҳга, жиноятга, разолатга бошловчи ёвуз рух ёки ёвуз рухларнинг бошлиғи, иблис (1); айёр, қув одам, алдамчи (2); шўх, ўйноқи (3).

<ноумид – шайтон> Инсон доим умид, яхши ният билан яшашини, умидсизланмасликни қайд этувчи, шунга даъват этувчи ибора.

<шайтон арава> эск. айн. велосипед

<шайтон йиғиси> Ёлғондакам, йўлига бўладиган йиғи.

<шайтонга ҳай бермоқ> Бирон хатти-ҳаракатдан ўзини тиймоқ, ўзини босмоқ.

<шайтони зўр> Шўхлиги тез ва кучли кўзғаладиган, шундай хусусиятли.

<шайтони кўзимок> Шайтони кўзғамок, жини тутмоқ.

Агар лексик бирлик контекстга қараб бир неча маънони англатса, контекстларда қўлланишига қараб керакли гуруҳга киритилади; ўша теглар ҳам битта мажмуага киради ҳамда бошқа теглардан нуқтали вергул билан ажратилади. Биринчи изоҳ асосий, нуқтали вергулдан кейинги изоҳ эса контекстдаги маъноси ҳисобланади. Разметкалаш ҳам сўз туркуми, ҳам алифбо тартибида амалга оширилади. Биринчи ёндашув бирликларнинг лексик-грамматик хусусиятини ёритиш – қайси туркумга мансублигини билдириш учун қўлланилади.

Семантик разметка тизимида тег гуруҳлари. Қалмиқ тили корпуси тажрибаси кўрсатдики, семантик разметка морфологик разметка асосида бажарилади: морфологик белгилар жадвалига яна тўртта устун қўшилган бўлиб, бу устунлар сўзшаклнинг семантик хусусиятини ифодалайди. Қуйида семантик разметка мажмуини қалмиқ тили корпуси мисолида кўриб чиқамиз.

1. Лексик-грамматик разметка: сўз туркумига тегишлилик белгисини ички разрядларга бўлиш. Бу таснифни амалга оширишда тил грамматикасида мавжуд лексик-семантик классификациядан фойдаланиш мумкин. Корпус муаллифлари, одатда, теглар мажмуини инглиз тилидаги қисқартма ёки шартли белгилар билан шаклантиришади[5].

От сўз туркуми морфологик белгисига қўшимча семантик теглар:

1. r : p r o r e r – атоқли от;

1.1. n a m e – исм;

1.2. s u r n – фамилия;

1.3. p a t r – отчество;

1.4. g e o x – топоним;

1.5. o r g n – ташкилот;

2. r : a r p r e l – жамловчи от;

2.1. c o n c r – нарса номи;

2.2. a b s t r – мавҳум от.

Сифат сўз туркуми морфологик белгисига қўшимча семантик теглар:

1. t : q u a l – аслий сифат;

2. t : r e l – нисбий сифат.

Сон сўз туркуми морфологик белгисига қўшимча семантик теглар:

1. o r d – тартиб сон;

2. c a r d – миқдор сон;

3. p a r – тақсим сон;

4. a g e – ёшни билдирувчи сон;

5. c o l – жамловчи сон.

Олмош сўз туркуми морфологик белгисига қўшимча семантик теглар:

1. p e r s – кишилик олмоши;

2. d e m – кўрсатиш олмоши;

3. q u a – белгилаш олмоши;

4. r e f l – ўзлик олмоши;
 5. i n t e r – сўроқ олмоши;
 6. i n d – гумон олмоши.

II. Лексик-тематик разметка: лексик-семантик гуруҳ, тўда аниқланади; бирликлар таксономик, мереологик, топологик муносабатда жойлаштирилади. В.В. Куканованинг фикрича, сўз семантикаси тил константаси саналади. Масалан, бир семантик гуруҳга турли туркумга мансуб бирликлар кириши мумкин.

III. Услубий баҳо характеристикаси: коннотацияни аниқлаш. Сўзнинг ижобий ёки салбий (агар мавжуд бўлса) оттенка билдириши аниқланади (ev: posit; e v: neg).

IV. Деривацион характеристика: агар ясалиш ҳолати мавжуд бўлса, сўзшаклнинг ўзак-негизи, морфо-семантик сўз ясалиш гуруҳи аниқланади. Сўзнинг лексик-семантик хусусияти сўз ясалиш кўрсаткичлари билан чамбарчас боғлиқ. Масалан, сўз –лик кўшимчаси билан тугаса, от/сифат туркумига, -чи билан тугаган сўз шахс билдирувчи сўзлар гуруҳига мансуб бўлади.

Қалмиқ тилида ҳайвонлар ёши, жинсига кўра фарқланади, номланиши ҳам шундай хусусиятга эга. Шундан келиб чиққан ҳолда, ҳайвон номлари тегларига кўшимча равишда t: animal: age, t: animal: gender; икки белги қўшилганда эса t: animal: age& gender теглари бириктирилади.

Барча сўз туркумларига тегишли бўлмаган таксонлар – бўлакни бирлаштирувчи *бутун* номи теглар тизимига киритилмаган: барча бирликларга бирдек тааллуқли семантик майдон номларигина фойдаланувчи интерфейсида кўринади.

Семантик теглашда энг асосий муаммо маълум бирликнинг контекстга кўра икки ва ундан ортиқ семантик гуруҳ/майдонга мансублик ҳолатини аниқлашдир. Сўз туркумлари ичидаги лексик бирликни тизимлаштириш шаклдан маънога қараб эмас, балки маънодан шаклга принципига ёндашуви асосида амалга оширилади: туркум ичидаги ҳар бир сўз алоҳида изоҳланмайди (бундай изоҳлаш морфологик аннотациялашга хос), балки семантик майдон ёки гуруҳ/тўда аниқланиб, шу гуруҳга мансуб сўзлар йиғилади. Бундай ёндашув тил бирлигига тафаккур ва мулоқот бирлиги сифатида қарашдан келиб чиқади.

А.А. Кретов А.Е. Кибрикнинг фикрига таянган ҳолда шундай хулосага келади: “Лексик ва грамматик семантика бир бутунлик ҳам эмас, алоҳида ҳам эмас. Лексик семантика – константа, грамматик семантика – ўзгарувчан. Лексик семантика бирламчи, грамматик семантика – ҳосилавий ҳодиса. Рус тили миллий корпуси семантик теглари тизими нисбатан оммалашган қарашни акс эттиради: семантик изоҳлар лексик ҳам эмас, грамматик ҳам эмас. Унинг асосий таксономи – сўз туркуми; у бирликнинг грамматик хусусиятини кўрсатиб туради. Предмет ва нопредмет отлар учун мереологияга тегишлилик (бутун-бўлак, элемент-кўприк), предмет отлари учун топология, феъллар учун куазация, предмет ва нопредмет отлар, сифат ва равиш учун баҳо оттенкаси каби семантик изоҳлар билан тўлдирилади”[4]. Корпус (НКРЯ) семантик разметка тизими муаллифларининг фикрича, ҳар бир сўз туркуми семантик изоҳлар тизими ўзига хос. Ҳатто ҳар бир разряд: нарса, шахс, ўрин-жой отлари, сифатнинг семантик разметка тизими ҳам фарқланади[8].

Бир сўз туркуми семантик теглар тизими бошқа туркумни изоҳлай олиш/олмаслигини асосий изоҳлар тизими мисолида кўриб ўтамиз.

Sn p t: action фаолият, жараён, тадбир (аукцион, кеча, сайлов, туғилган кун, мажлис, йиғилиш, саёҳат) *ўзгарувчан hum.*

Sp t: animal ҳайвонот (сигир, жирафа, қарға, эчкемар, чумоли) *константа.*

Sn p t: color ранг (ранг-тус, колорит, сарғайиш, яшариш) *ўзгарувчан.*

Sp t: constr бино ва иншоот (уй, чордоқ, кўприк) *ўзгарувчан.*

Sn p t: disease касаллик (ангина, диабет) *ўзгарувчан.*

Ss t: famn фамилия (Пушкин) *ўзгарувчан.*

Sp t: food овқат ва ичимлик (пирог, бўтқа, сут) *ўзгарувчан.*

Sn p t: game ўйин (покер, домино, волейбол) *ўзгарувчан.*

Sp t:inter ўзаро таъсир ва муносабат (ўзаро ёрдам, уруш, тўлғоқ, жанжал) ўзгарувчан hum.

Sp t: param ўлчам (баландлик, оғирлик) ўзгарувчан.

Ss t: patrн отчество (Шарипович) ўзгарувчан.

Ss t: persн исм (Алишер) ўзгарувчан.

Sp t: plant ўсимлик (райхон, атиргул, ўт) – константа.

Sp t: space ўрин-жой (космос, шаҳар, тайга, ўрмон, чакалакзор, манзар). Булардан айрими инсон аралашуви, айрими табиий ҳолатда яратилган, шунинг учун ўзгарувчан ва константа.

Sp t: sport спорт (спартакиада, акробатика, баскетбол) ўзгарувчан.

Sp t: stuff нарса ва материал (сув, кум, тупроқ, хамир, ипак) Булардан айрими инсон аралашуви, айрими табиий ҳолатда яратилган, шунинг учун ўзгарувчан ва константа.

Sp t: taste маза-таъм (мазали, аччиқ, нордон)

Sp t: sur'ater температура (салқин, илиқ, иссиқ) ўзгарувчан.

Sp t: text матн (ҳикоя, китоб, эълон) ўзгарувчан.

Ss t: topon топоним (Европа, Волга, Эльбрус, Ҳимолай, Мовароуннаҳр) t:natur:space:topon (табиий топоним); Бухоро, Тошкент t: hum: space: topon (инсоният томонидан яратилган топоним) ўзгарувчан ва константа.

Sp t: unit ўлчов бирликлари (балл, килограмм, метр, дақиқа)

Адекват t:time теги билан бирга ушбу изоҳни аниқлаштирувчи t:time:age (ёш) ҳамда t:time:age:abs теглари ҳам ишлатилади. Бунда t:time теги доимий, ўзгармас вақт бирликларини билдирса, қуйидаги изоҳлар ушбу изоҳни яна ҳам аниқлаштириб кўрсатади:

A t:time:age:abs абсолют (уч яшар)

A t:time:age:max катта (тўнғич, қари, қадимий)

A t:time:age:min кичик (ёш, кичик)

A t:time:age: ёш (қари, ёши улуғ)

Sp t:time:age: ёш (болалик, ёшлиқ, йигирма йиллик).

Ишнинг аввалги қисмида семантик разметка тизимида константа ва оператор-классификатор типдаги тег (изоҳ)лар кириши ҳақида айтиб ўтилганди. Юқорида мисол сифатида келтирилган семантик разметка тизимида теглар ўзгарувчан ва константа сифатида кўрсатилган. Лексик бирлик контекстан қатъи назар битта маънони англатса, константа белгиси, контекстга қараб маъноси ўзгарадиган бўлса, ўзгарувчанлик белгиси қўйилади. Шунингдек, табиий топонимларга константалик белгиси, инсон томонидан барпо этилган топонимларга ўзгарувчанлик белгиси қўйилган; бошқа ҳолатда ҳам инсон томонидан содир/барпо этилган нарса/ҳодиса/предмет/жой шундай белгиланади.

А.А. Кретов НКРЯ[15] семантик разметка тизимини таҳлил қилар экан, қуйидагиларни ўта аниқ ишланган теглар тизими сифатида кўрсатади:

Sp t:hum:etn – этноним (эфиоп, итальян);

Sp t:hum:kin – қариндошлиқни билдирувчи номлар (ака, буви, амма);

Sp t:hum:supernat – ғайритабиий мавжудотлар (сув париси, ўзга сайёралик)

t:persн – исм (Алишер);

t:patrн – отчество (Шарипович);

t:famn – фамилия (Маҳмудов);

t:topon – топоним (Европа, Волга, Эльбрус, Москва, Мовароуннаҳр, Тошкент).

Ушбу аниқ теглар омоним ва кўп маъноли бирликларни теглашда жуда муҳим, чунки бу теглар бир маъноли бирликларни изоҳлайди. Шунингдек, ўрин-жойни билдирувчи изоҳлар place, loc и space шаклига эга. А.А. Кретов бу тегларни ишлатишни оқламайди.

ADV t:place – ўрин (шу ердаги, ўртадаги);

A t:place – ўрин (ўнг, чап, чет);

V t:loc – жойлашув (ётмоқ, турмоқ, қўймоқ);

Sp t:loc – жойлашув (жойлашган ўрин);

Sp t:space – бўшлиқ ва жой (космос, шаҳар, тайга, овлок, даҳлиз).

Ўрин-жойни билдирган ушбу теглар бошида бирликнинг қайси туркумга тегишлилигини билдирувчи белги мавжуд: ADV сифатни, A от, V феъл, Snp ва Sp ўрин-жойни билдириши билан ажралиб туради. Биз А.А.Кретовнинг фикрига қўшилмаймиз. Бизнингча, ушбу теглар ҳам изоҳни конкретлаштиришга хизмат қилади ва уларни синтактик разметка тизимига киритиш ўринли.

Қуйида санаб ўтиладиган теглар сифат туркумига хос. Кўринадикки, от сўз туркумига мансуб сўзларни бу изоҳлар билан теглаб бўлмайди:

t:color – ранг (бўёқ, колорит, сарғайиш, кўкариш)

A t:physq:color ранг (қизил, туссиз);

A t:physq:form шакл (эгри, айлана);

A t:physq:smell хид (хушбўй, муаттар, сассик);

A t:physq:taste маза-таъм (нордон, чучук);

A t:physq:sur'ater температура (иссик, муздек);

A t:physq:weight оғирлик (оғир, енгил);

A t:physq физик хусусият (юмшок, сикик).

Корпус семантикасида “Бутун-бўлак” майдонида алоҳида ўрин ажратилган; интерфейсида ҳам алоҳида ойнада жойлашади. “Бутун-бўлак” майдонида инсон тана аъзолари: қулоқ, бурун, оғиз, лаб, қўл, оёқ, бармоқ, кўз, қош кабиларни киритиш лисоний мантикка мос тушади.

А.А. Кретов НКРЯ семантик теглар тизимида 80 та тегни ажратиб кўрсатади, улардан 6 таси икки позицияда учрайди, 74 таси қайта учрамайдиган теглардир. Масалан:

Snp t:be:appear мавжудликнинг бошланиши (пайдо бўлиш, туғилиш, шаклланиш, ташкилот, вужудга келиш);

V t:be:appear мавжудликнинг бошланиши (пайдо бўлмоқ, туғилмоқ, шаклланмоқ, вужудга келмоқ);

Snp t:be:disapp мавжудликнинг тугаши (ўлим, қатл, тугатиш/ёпиш);

V t:be:disapp мавжудликнинг тугаши (ўлмоқ, ўлдирмоқ, тугатмоқ);

Snp t:be:exist мавжудлик (ҳаёт, мавжудлик);

V t:be:exist мавжудлик (яшамоқ, содир бўлмоқ);

Snp t:be: маиший соҳа (яшаш, пайдо бўлиш, ўлим);

V t:be: маиший соҳа (яшамоқ, пайдо бўлмоқ, ўлдирмоқ).

Кўриниб турибдики, ушбу ўхшаш теглар бошқа-бошқа туркумларни изоҳлайди, фақат белгиси такрор келган, холос. Қуйида кўрсатилган изоҳлар орасидаги фарқни ҳам ажратиш қийин:

V t:be: маиший соҳа (яшамоқ, содир бўлмоқ);

V t:be:exist мавжудлик (яшамоқ, содир бўлмоқ);

V t:be:disapp мавжудликнинг тугаши (ўлмоқ, ўлдирмоқ, тугатмоқ);

V t:be:appear мавжудликнинг бошланиши (пайдо бўлмоқ, туғилмоқ, шаклланмоқ, вужудга келмоқ).

Фикримизча, ушбу ўхшаш, такрор келадиган теглар бир хил маъноли бирликларни изоҳлаганлиги сабабли улардан фақат биттасини (Масалан, V t:be:disapp ҳамда V t:be:appear) колдириш тўғри бўлади.

А.А. Кретов семантик разметка тизими таҳлили асосида хулоса қилар экан, қуйидагиларни алоҳида санаб ўтади[4]:

1. Сўз туркумлари орасида оралиқ учинчи мақомига эга бирликлар семантик тегларнинг аниқлигига соя солади.

2. Лексик ва грамматик семантика ажралмас белги/параметр бўлиши лозим. Бундай изоҳлашда константа ва классификатор бир-бирига тўғри келмай қоладиган ўринлар ҳам бўлади. Бу ҳолат семантик теглар тизимида номутаносибликни келтириб чиқаради.

3. Семантик классификациянинг асосий вазифаси константани аниқлашдан иборат.

4. Семантик константа мажмуи тилдаги ўзак – туб сўзлар мажмуидан келиб чиқади ҳамда улар дунёнинг архаик картинасини ташкил қилади.

5. Семантик теглар мажмуини тўлиқ санаб ўтишнинг имкони бўлмаса-да, уларнинг энг асосийлари сифатида табиат, инсон, ҳайвонот, ўсимликлар олами, предмет каби белгиларни санаб ўтиш жоиз.

Хулоса сифатида таъкидлаш жоизки, семантик разметка тизими умумлингвистик разметка тизими хусусиятлари билан бирга бир қанча хусусий белгиларга ҳам эга. Бошқа лингвистик разметка тизимларига ўхшаш томони разметканинг лингвистик таъминот ёки базага асосланишидир. Семантик разметканинг ўзига хос томони аннотациялашнинг икки хил: фасет ва шажара усулининг мавжудлигида. Амалда бирликнинг хусусиятига кўра бу икки усулдан бири ёки иккаласи аралаш ҳолда қўлланади. Фасет усули бирликка кетма-кет изоҳ беришни тақозо қилса, шажара усули бирлик мансуб бўлган гуруҳ, тўда, майдонни белгилашни талаб қилади.

Биринчи усулда сўздан маънога, иккинчи усулда маънодан сўзга қараб борилади. Бу мутахассиснинг разметка жараёнида ҳам, интерфейсида акс этадиган қидирувда ҳам сўз → маъно, маъно → сўз принципининг амал қилишини кўрсатади. Мавжуд online корпусларни кузатиш натижасида разметка тизимида лексик-грамматик; лексик-тематик; услубий баҳо характеристикаси; деривацион характеристика каби белгиларнинг устуворлигига гувоҳ бўлдик. Константа ва оператор-классификатор теглар ҳам разметка тизимидан доимий жой оладиган тег ҳисобланади.

АДАБИЁТЛАР

14. Бусқунбаева Л.А., Сиразетдинов З.А. Система разметок в Национальном корпусе башкирского языка [электронный ресурс] // URL: <http://www.marlamuter.ru/forum/index.php?topic=21>.

15. Загоруйко 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>

16. Караулов Ю.Н., Молчанов В.И., Афанасьев В.А., Михалев Н.В. Русский семантический словарь: Опыт автоматического построения тезауруса: от понятия к слову / отв. ред. С.Г.Бархударов. – М.: Наука, 1983. – 566 с.

17. Кретов А.А. Анализ семантических помет в НКРЯ // <http://ruscorpora.ru/sbornik2008/11.pdf>

3. Куканова В.В. Принципы семантической разметки национального корпуса калмыцкого языка // http://kalmcorp.ru/sites/default/files/kukanova_25.pdf

4. 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. – М.: Индрик, 2005. – С. 155-174.

5. Кустова Г.И., Толдова С.Ю. НКРЯ: семантические фильтры для разрешения многозначности глаголов. // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы / отв. ред. В.А.Плунгян. – СПб.: Нестор-История, 2009. – С. 258-278.

6. Ляшевская О.Н. Топологические классы имен в семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). – СПб.: С.-Петербургский гос.университет, Факультет филологии и искусств, 2008 – С. 276-284.

7. Ляшевская О.Н. Топологические классы имен в семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). – СПб.: С.-Петербургский гос.университет, Факультет филологии и искусств, 2008. – С. 276-284.

8. Ляшевская О.Н. Топологические классы имен в семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). – СПб.: С.-Петербургский гос.университет, Факультет филологии и искусств, 2008. – С. 276-284.

9. Рахилина Е.В., Кобрицов Б.П., Кустова Г.И., Ляшевская О.Н., Шеманаева О.Ю. Многозначность как прикладная проблема: семантическая разметка в Национальном корпусе русского языка. // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог-2006». – М., 2006. – С. 445-450.

10. Рахилина Е.В., Кустова Г.И., Ляшевская О.Н., Резникова Т.И., Шеманова О.Ю. Задачи и принципы семантической разметки лексики в НКРЯ // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы / отв. ред. В. А. Плунгян. – СПб.: Не-стор-История, 2009. – С. 215-239.

11. Русский семантический словарь. Толковый словарь, систематизированный по классам слов и значений. / Ин-т рус. яз. им. В. В. Виноградова РАН; под общ. ред. Н. Ю. Шведовой. Т. 1. – М.: Азбуковник, 1998. XXV, 807с. Т. 2. – М.: Азбуковник, 2000. XXXII, 762 с. Т. 3. М.: Азбуковник, 2003. 720 с. Т. 4. – М.: ИРЯ РАН, 2007. – 952 с.

УДК: 4И(Ам)075ТЭ78

ҲАРБИЙ-ПРОФЕССИОНАЛ ТИЛ БИРЛИКЛАРИНИНГ СЎЗ ТУЗУЛМАСИДАГИ СЕМАНТИК-СИСТЕМ ТАВСИФИ

СЕМАНТИКО-СИСТЕМНАЯ ХАРАКТЕРИСТИКА ВОЕННО- ПРОФЕССИОНАЛЬНЫХ ЕДИНИЦ СВЕРХСЛОВНОЙ СТРУКТУРЫ

DESCRIPTION OF THE SEMANTIC SYSTEM IN THE WORD FORMATION OF MILITARY-PROFESSIONAL LANGUAGE UNIT

Эркинов Сухроб Эркинович

*зав. каф. иностранных языков, теории и
практики английского языка, СамГИ*

Аннотация. *Мазкур мақола ҳарбий-профессионал тил бирликларининг сўз тузулмасидаги тизимли тавсифига бағишланган бўлиб, унда инглиз тилида сўзлашувчи ҳарбий соҳада қайд этилган икки компонентли номинал ибораларнинг семантик хусусиятларини ёритиб беришга қаратилган.*

Таянч сўзлар: *сўз бирикмаси, термин, ҳарбий соҳа, лексик бирлик, тизим, семантик майдон, сўз парадигмаси, тил структураси.*

Аннотация. *Статья посвящается системному описанию словесной структуры военно-профессиональных языковых единиц, основное внимание направлено на выделение семантических особенностей двухкомпонентных номинальных выражений в англоговорящей военной сфере.*

Ключевые слова: *словосочетание, термин, военная сфера, лексическая единица, системность, семантическое поле, парадигма слов, языковая структура.*

Annotation. *This article is devoted to the systematic description of military-professional language units in the word structure, which aims to highlight the semantic features of two-component nominal expressions in the English-speaking military sphere.*

Key words: *phrase, term, military sphere, lexical unit, consistency, semantic field, word paradigm, language structure.*

Введение. В современном английском языке выделяются своей деривационной продуктивностью именные словосочетания с препозитивными определителями. Особо важную роль они играют в процессе формирования и закрепления понятий, которые относятся к различным областям знаний (Митчелл, 2014: 69-70; Дониева, 2017). Достаточно отметить, что в разделе новых слов [Addenda] словаря Уэбстера [Webster - 2008] на долю подобных словосочетаний, рассматриваемых как самостоятельные единицы языка, приходится порядка 20 % его состава.

Так, по данным Э. Бэтса (Bates, 2006: 201-229), из общего количества [24 000] именных словосочетаний с препозитивными определениями, зафиксированных при словарном обследовании специальных слов и выражений военной тематики, двухкомпонентных словосочетаний оказалось порядка 6 000, а словосочетаний с тремя и более компонентами – около 2 000.

Роль ядра в таких сочетаниях играют существительное, а в качестве определяющих слов чаще всего выступают существительные [task force], прилагательные [basic policy] и причастия [delayed action]. Проблема разграничения словосочетания и сложного слова нами не затрагивается. Все именные комплексы слов, разделенных на письме пробелом, относятся к категории словосочетаний.