

**БУХОРО ДАВЛАТ УНИВЕРСИТЕТИ**  
**ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ**  
**PhD.03/27.02.2020.Fil.72.09 РАҚАМЛИ ИЛМИЙ КЕНГАШ**

---

**БУХОРО ДАВЛАТ УНИВЕРСИТЕТИ**

**АХМЕДОВА ДИЛДОРА БАҲОДИРОВА**

**АТОВ БИРЛИКЛАРИНИ ЎЗБЕК ТИЛИ КОРПУСЛАРИ УЧУН**  
**ЛЕКСИК-СЕМАНТИК ТЕГЛАШНИНГ ЛИНГВИСТИК АСОС ВА**  
**МОДЕЛЛАРИ**

**10.00.01 – Ўзбек тили**

**Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси**  
**АВТОРЕФЕРАТИ**

**Бухоро – 2020**

УДК: 81'1:81'37[=512.133]

**Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси  
автореферати мундарижаси  
Оглавление автореферата диссертации доктора философии (PhD) по  
филологическим наукам  
Contents of dissertation abstract of doctor of philosophy (PhD)  
in philological Science**

**Ахмедова Дилдора Баходировна**

Атов бирликларини ўзбек тили корпуслари учун лексик-семантик  
теглашнинг лингвистик асос ва моделлари.....5

**Ахмедова Дилдора Баходировна**

Лингвистические основы и модели лексико-семантической разметки  
номинативных единиц для корпусов узбекского языка..... 26

**Axmedova Dildora Bakhodirovna**

Linguistic bases and models of lexical-semantic tagging of nominative units for  
Uzbek language corpora.....49

**Эълон қилинган ишлар рўйхати..... 53**

Список опубликованных работ

List of publications

**БУХОРО ДАВЛАТ УНИВЕРСИТЕТИ**  
**ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ**  
**PhD.03/27.02.2020.Fil.72.09 РАҚАМЛИ ИЛМИЙ КЕНГАШ**  

---

**БУХОРО ДАВЛАТ УНИВЕРСИТЕТИ**

**АХМЕДОВА ДИЛДОРА БАҲОДИРОВА**

**АТОВ БИРЛИКЛАРИНИ ЎЗБЕК ТИЛИ КОРПУСЛАРИ УЧУН**  
**ЛЕКСИК-СЕМАНТИК ТЕГЛАШНИНГ ЛИНГВИСТИК**  
**АСОС ВА МОДЕЛЛАРИ**

**10.00.01 – Ўзбек тили**

**Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси**  
**АВТОРЕФЕРАТИ**

**Бухоро–2020**

Фалсафа доктори (PhD) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида № В 2019.2.PhD/Fil 777 рақам билан рўйхатга олинган.

Диссертация Бухоро давлат университетида бажарилган.

Диссертация автореферати уч тилда (ўзбек, рус, инглиз (резюме) Илмий кенгашнинг веб-саҳифаси [www.buxdu.uz](http://www.buxdu.uz) ва «Ziyonet» Ахборот таълим портали ([www.ziyonet.uz](http://www.ziyonet.uz))да жойлаштирилган.

**Илмий раҳбар:** Менглиев Бахтиёр Ражабович,  
филология фанлари доктори, профессор

**Расмий оппонентлар:** Раупова Лайло Рахимовна,  
филология фанлари доктори, профессор  
Абжалова Манзура Абдурашатовна,  
филология фанлари бўйича фалсафа доктори (PhD)


**Етакчи ташкилот:** Қарши давлат университети

Диссертация ҳимояси Бухоро давлат университети ҳузуридаги PhD.03/27.02.2020.Fil.72.09 рақамли Илмий кенгашнинг 20 20 йил 20 октябрь соат 10<sup>00</sup> даги мажлисида бўлиб ўтди. (Манзил: 200118, Бухоро шаҳри, М. Иқбол кўчаси, 11-уй. Тел.: (0 365) 221-29-14; факс: (0 365) 221-29-14, email: [buxdu\\_rektor@buxdu.uz](mailto:buxdu_rektor@buxdu.uz)).


Диссертация билан Бухоро давлат университети Ахборот-ресурс марказида танишиш мумкин (968 рақами билан рўйхатга олинган). (Манзил: 200118, Бухоро шаҳри, М. Иқбол кўчаси, 11-уй. Тел.: (0 365) 221-29-14).

Диссертация автореферати 20 20 йил « 8 » октябрь да тарқатилди.  
(20 20 йил « 7 » октябрь даги 1-рақамли реестр баённомаси).



  
М.К.Абузалова  
Илмий даражалар берувчи Илмий кенгаш раиси, филология фанлари доктори (DSc), профессор

  
Г.И.Тоирова  
Илмий даражалар берувчи Илмий кенгаш илмий котиби, филология фанлари бўйича фалсафа доктори (PhD), доцент

  
Д.З.Ражабов  
Илмий даражалар берувчи Илмий кенгаш қошидаги Илмий семинар раиси, филология фанлари доктори (DSc), доцент

## КИРИШ (фалсафа доктори (PhD) диссертацияси аннотацияси)

**Диссертация мавзусининг долзарблиги ва зарурати.** Жаҳон тилшунослигида компьютер ва корпус лингвистикаси муаммоларини ўрганиш жараёни XX асрнинг 40-йилларида бошланиб, 60-йилларда жадаллашди, XXI аср бошларида миллионлаб сўзлардан иборат тил корпуслари пайдо бўлди. Автоматик таржима, тезаурус, электрон луғатлардан фойдаланиш имконияти кенгайди, илмий-назарий асослари яратилди, амалиётда қўллана бошлади. Мазкур янгиланишлар ахборот технологияларини тилшуносликка татбиқ этиш билан боғлиқ истиқболли илмий йўналишлар пайдо бўлишига йўл очди. Бу эса атов бирликларини корпус материали сифатида теглаш тамойилларини ўрганиш заруратини кун тартибига қўйди.

Дунё тилшунослигида XXI асрга келиб, корпус лингвистикасини илмий-назарий жиҳатдан ўрганиш ҳаракати кучайди. Компьютер лингвистикаси йўналишида автоматик таржима сифатини яхшилаш, ҳар бир тилга оид сўзларни леммалаш ва теглаш назарияси, алгоритми ҳамда лингвистик таъминотини яратиш, матн таҳлил қилиш, теглаш, таржима қилиш дастурлари тузиш жаҳон тилшунослигида долзарб масалага айланди. Тилшуносликда, хусусан, компьютер лингвистикаси соҳасида корпус бирликларини семантик теглаш, тег моделлари, теглаш алгоритмини ишлаб чиқиш, шу асосда автоматик семантик теглаш дастурини ишлаб чиқишга зарурат сезилмоқда.

Истиқлол йилларида компьютер лингвистикасида автоматик таржима, сунъий интеллектнинг ўзбек тилини тушуниш ва қайта ишлашига эришиш борасида талай ишлар амалга оширилган. Бинобарин, барча илмий йўналишлар қаторида тилшуносликда ҳам «...Давлат тилини ривожлантиришга оид илмий-тадқиқот ишларини қўллаб-қувватлаш, бу соҳада халқаро ҳамкорликни амалга ошириш вазифаси» нинг<sup>1</sup> белгиланиши ўзбек тилининг интернет тизимидан мустаҳкам ўрин олишига кўмаклашиш билан боғлиқ изланишлар олиб бориш муҳимлигини кўрсатади. Мамлакатимизда давлат тилига эътибор маънавиятга эътиборнинг устувор йўналишларидан бири даражасига кўтарилди. Шу боисдан она тилимизни асраб-авайлаш, бойитиш, ундан амалий фойдаланиш самарадорлигини ошириш билан бирга ўзбек тилининг замонавий ахборот-коммуникация тизимида кенг қўлланишига эришиш кечиктириб бўлмас масалага айланди. Корпус лингвистикаси истиқболли илмий йўналиш эканлигини инобатга олган ҳолда ўзбек тили миллий корпусини яратиш, лингвистик моделларни тузиш сингари масалаларни замонавий илмий тамойиллар асосида тадқиқ этиш фанимиз олдида турган долзарб вазифа ҳисобланади.

Ўзбекистон Республикаси Президентининг 2016 йил 13 майдаги «Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетини ташкил этиш тўғрисида»ги ПФ-4997-сон, 2017 йил 7

---

<sup>1</sup>Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг «Ўзбек тилининг давлат тили сифатидаги нуфузи ва мавқеини тубдан ошириш чора-тадбирлари тўғрисида»ги фармони // [www.xabar.uz](http://www.xabar.uz)

февралдаги «Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида»ги ПФ-4947-сон Фармонлари, 2017 йил 17 февралдаги «Фанлар академияси фаолияти, илмий тадқиқот ишларини ташкил этиш, бошқариш ва молиялаштиришни янада такомиллаштириш чора-тадбирлари тўғрисида»ги ПҚ-2789-сон Қарори, Ўзбекистон Республикаси Президентининг 2019 йил 21 октябрдаги «Ўзбек тилининг давлат тили сифатидаги нуфузи ва мавқеини тубдан ошириш чора-тадбирлари тўғрисида»ги ПФ-5850-сон Фармони, Ўзбекистон Республикаси Вазирлар Маҳкамасининг 2019 йил 12 декабрдаги «Давлат тилини ривожлантириш департаменти тўғрисидаги Низомни тасдиқлаш ҳақида»ги 984-сон Қарори ҳамда мазкур фаолиятга тегишли бошқа меъёрий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишда ушбу диссертацион иш муайян даражада хизмат қилади.

**Тадқиқотнинг республика фан ва технологиялари ривожланишининг устувор йўналишларига мослиги.** Диссертация республика фан ва технологиялари ривожланишининг I. «Ахборотлашган жамият ва демократик давлатни ижтимоий, ҳуқуқий, иқтисодий, маданий, маънавий-маърифий ривожлантиришда инновацион ғоялар тизимини шакллантириш ва уларни амалга ошириш йўллари» устувор йўналишига мувофиқ бажарилган.

**Муаммонинг ўрганилганлик даражаси.** Жаҳон тилшунослигида корпус соҳасидаги мақсадли тадқиқотлар XX асрнинг 40-йилларида Блумфилд, Фриз ва Бонджерслар томонидан бошланган<sup>2</sup>; Н.Френсис ва Г.Кучера илк марта корпус тузиш принципларини ишлаб чиққан<sup>3</sup>. Рус тилшунослигида В.П.Захаров, А.Б.Кутузов, Е.В.Недошивина, В.В.Риков, В.А.Плунгянлар корпус, унинг турлари, корпус тузиш ва теглаш тамойиллари борасида тадқиқот олиб боришган<sup>4</sup>. Корпус бирликларини лексик-семантик теглаш муаммолари Г.И.Кустова, О.Н.Ляшевская, Е.В.Падучева, Е.В.Рахилина<sup>5</sup>, Ю.Д.Апресян, Л.Л.Иомдин, А.В.Санников, В.Г.Сизов тадқиқотлари<sup>6</sup> предмети бўлган.

<sup>2</sup>Блумфильд Л. Язык. – Москва: Прогресс, 1968. – 608 с.; Fries Ch.C. The structure of English. An introduction to the construction of English sentences. – New York: Harcourt, 1952. – 304 p.; Bongers H. The history and principles of Vocabulary control. – Woerden: WOCOPI, 1947. . – 442 p.

<sup>3</sup>Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – Москва, 1967.; Синклер Д. Предисловие к книге “Как использовать корпуса в преподавании иностранного языка”/ Д. Синклер [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/corpora-infro.html>, свободный.

<sup>4</sup> Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – 69 с. – С. 27.; Кутузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) – //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf; Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург, 2006.–26 с.; Рыков В.В. Курс лекций по корпусной лингвистике. URL: <http://rykov-cl.narod.ru/c.html>; Плунгян В. Зачем мы делаем Национальный корпус русского языка? “Отечественные записки” 2005, №2. [http://magazines.russ.ru/oz/2005/2/2005\\_2\\_20-pr.html](http://magazines.russ.ru/oz/2005/2/2005_2_20-pr.html)

<sup>5</sup>Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Семантическая разметка лексики в национальном корпусе русского языка: принципы, проблемы, перспективы // (Электрон ресурс): <http://ruscorpora.ru/sbornik2005/10kustova.pdf>;

<sup>6</sup>Апресян Ю.Д., Иомдин Л.Л., Санников А.В., Сизов В.Г. Семантическая разметка в глубоко аннотированном корпусе русского языка // Труды международной конференции “Корпусная

Ўзбек тилшунослигида компьютер лингвистикаси, матнга лексикографик ишлов бериш ва лингвостатистик таҳлил этиш борасида муайян тадқиқотлар амалга оширилган. А.Қ.Пўлатов, С.М.Муҳаммедов, Н.К.Айимбетов, С.Муҳамедова, С.А.Каримов, Г.У.Жуманазарова, А.Бабанаров, Д.Б.Ўринбоева, Н.З.Абдурахмонова, А.М.Норов ва бошқаларнинг изланишларини ана шундай ишлар сифатида қайд этиш ўринли<sup>7</sup>. Бу тадқиқотлар матни инновацион ёндашув – компьютер лингвистикаси ютуқлари ёрдамида лексикографик ва лингвостатистик тадқиқ этишнинг замонавий усуллари хусусида зарурий тавсиялар бера олганлиги билан долзарблик касб этган. Ўзбек тилшунослигида корпус, унинг ўзига хос хусусияти, ижтимоий, лексикологик, таълимий аҳамияти, корпус турлари, муаллифлик корпусининг лингвистик қиммати, ўзбек тили муаллифлик корпусини яратишнинг лингвистик асослари масалалари Ш.М.Хамроева<sup>8</sup>, ўзбек тили корпуслари учун синоним сўзлар базаси А.А.Эшмуминов<sup>9</sup> томонидан тадқиқ этилган. Диссертацияни тайёрлаш жараёнида юқорида номлари санаб ўтилган тилшуносларнинг тадқиқотлари атрофлича ўрганилди, муносабат билдирилди ва улардан тадқиқотда фойдаланилди.

**Диссертация тадқиқотининг диссертация бажарилган олий таълим муассасасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги.** Диссертация Бухоро давлат университетининг илмий тадқиқот режаларига мувофиқ «Ўзбек тилшунослигида тил, шахс ва жамият муносабатлари тадқиқи муаммолари» мавзусидаги илмий тадқиқот йўналиши доирасида бажарилган.

**Тадқиқотнинг мақсади** ўзбек тили атов бирликлари орасидаги лексик-семантик муносабатларнинг корпус тузишдаги аҳамиятини ўрганиш; атов бирликларини семантик теглашнинг моделлари ҳамда ўзбек тили

---

лингвистика – 2004”. – Санкт-Петербург: Издательство Санкт-Петербургского университета, 2004. – С.41-54.

<sup>7</sup>Муҳаммедов С.М. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дисс... канд. фил. наук.–Ташкент, 1980. – 88 с.; Бабанаров А. Разработка принципов построения словарного обеспечения турецко-русского машинного перевода: Автореф. дисс... канд. фил. наук. – Ленинград, 1981. –27 с.; Айимбетов Н.К. Опыт лингвостатистического анализа лексики и морфологии каракалпакского публицистического текста: Автореф. дисс... канд. фил. наук.– Ташкент, 1987.; Каримов С., Қаршиев А., Исроилова Г. Абдулла Қаҳҳор асарлари тилининг луғати. Алфавитли луғат. Частотали луғат. Терс луғат. – Тошкент, 2007. – 420 б.; Ризаев С. Ўзбек тилининг лингвостатистик тадқиқи: Фил.фан.док.дис...автореф. – Тошкент, 2008. – 46 б.; Муҳаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. – Тошкент, 2006. – 45 б.; Ўринбоева Д.Б. Ўзбек фольклори матнларининг лингвостатистик тадқиқи. – Тошкент: Фан, 2010. – 121 б.; Жуманазарова Г.У. Фозил Йўлдош ўғли дostonлари тилининг лингвопоэтикаси: Фил. фан. док. дис...автореф. – Тошкент, 2017. –56 б. ; Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (содда гаплар мисолида): Фил.фан.бўйича фалсафа доктори (PhD)...дис. автореф. – Тошкент, 2018. – 55 б. ; Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011. – 175 б.; Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.

<sup>8</sup>Хамроева Ш.М. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Фил.фан.бўйича фалсафа доктори (PhD)...дис. – Бухоро, 2018. – 250 б.

<sup>9</sup>Эшмуминов А.А. Ўзбек тили миллий корпуси синоним сўзлар базасини яратиш асослари: Фил.фан.бўйича фалсафа доктори (PhD)...дис. – Термиз, 2019. – 140 б.

бирликларини лексик-семантик теглашнинг корпус учун лингвистик таъминотини ишлаб чиқишдан иборат.

**Тадқиқотнинг вазифалари:**

Ўзбек тилида лексик-семантик майдон назариясининг корпус тузишдаги аҳамияти ҳамда лексик-семантик муносабатларнинг атов бирликларини теглашдаги ўрнини тадқиқ этиш;

корпус лингвистикасида ўзбек лексикографияси ютуқларидан фойдаланиш имкониятларини аниқлаш;

лексик-семантик теглаш назариясидаги қарашларни ўрганиш;

семантик теглаш алгоритминини аниқлаш;

Ўзбек тили семантик майдонлари тег моделларини ишлаб чиқиш;

Ўзбек тилидаги омоним ва кўп маъноли сўзларни теглаш муаммоларини аниқлаш ҳамда шундай бирликларни теглашнинг оптимал усулларини ишлаб чиқиш;

Абдулла Қаҳҳорнинг «Анор» ҳикояси матнинини семантик теглаш ва натижаларни умумлаштириш.

**Тадқиқотнинг объекти** сифатида ўзбек тили атов бирликлари танланган.

**Тадқиқотнинг предмети** лексик-семантик теглашнинг лингвистик асос ва моделлари ташкил этади.

**Тадқиқотнинг усуллари.** Диссертация мавзусини ёритишда таснифлаш, тавсифлаш, қиёслаш, статистик таҳлил методларидан фойдаланилган.

**Тадқиқотнинг илмий янгилиги** қуйидагилардан иборат:

луғавий-маъновий майдон назариясининг корпус тузишдаги аҳамияти ўрганилган, ўзбек тилидаги луғавий-маъновий муносабатларнинг ўзбек тили атов бирликларини маъновий теглашдаги ўрни асосланган;

луғатшунослик маҳсулотларининг тил корпусини семантик теглаш учун лингвистик таъминот вазифасини бажариши исботланган;

Ўзбек тили атов бирликлари учун маъновий тег категориялари, константа ва оператор тег моделлари тузишнинг алгоритми ишлаб чиқилган;

Ўзбек тилидаги кўп маъноли сўзларни теглашнинг назарий асослари, уларни теглаш учун семантик филтър тузиш йўллари далилланган.

**Тадқиқотнинг амалий натижалари** қуйидагилардан иборат:

Ўзбек тили семантик майдонлари асосида семантик тег моделлари аниқланган;

Ўзбек тилидаги омонимларни теглаш муаммолари аниқланиб, ечимлар исботланган;

семантик тег категория ҳамда моделларини тузиш алгоритми ишлаб чиқилган;

Ўзбек тилидаги атов бирликларини семантик теглаш ҳамда лингвистик моделлаштириш усуллари далилланган;

Абдулла Қаҳҳорнинг «Анор» ҳикояси матнидаги атов бирликлари семантик тегланган ва натижалар амалиётда аниқланган.

**Тадқиқот натижаларининг ишончлилиги** ўрганилган материалларнинг ўзбек тили табиатидан келиб чиққан ҳолда хулосалар



килишга ёрдам берганлиги, уларнинг асосли эканлиги, методологик мукамаллиги, ўзбек тили атов бирликларини семантик теглаш тамойилларини яратишда амалда исботланган манбаларга таянилганлиги билан изоҳланади.

**Тадқиқот натижаларининг илмий ва амалий аҳамияти.** Тадқиқот ўзбек миллий корпуслари яратишнинг назарий асосларини ишлаб чиқишда, компьютер лингвистикаси йўналишида тадқиқотлар яратишда илмий-назарий манба сифатида хизмат қилади.

Тадқиқотнинг амалий аҳамияти амалий филология бўлимлари: компьютер лингвистикаси, корпус лингвистикасининг фан сифатида ўқитилиш жараёнида дастур, режа тузиш ҳамда мавзуларни баён этишда манба вазирасини ўташи билан изоҳланади.

**Тадқиқот натижаларининг жорий қилиниши.** Тадқиқотда илмий асосланган ўзбек тили атов бирликларини лексик-семантик теглаш, лексик-семантик тег моделлари тузиш алгоритми тавсифи асосида:

ўзбек тилида лексик-семантик майдон назариясининг корпус тузишдаги аҳамияти ҳамда лексик-семантик муносабатларнинг ўзбек тили атов бирликларини маъновий теглашдаги ўрнига доир хулосалардан 2012-2016 йилларда бажарилган Ф1-ХТ-0-19919-рақамли «Ўзбек мифологияси ва унинг бадиий тафаккур тараққиётида тутган ўрни» мавзусидаги фундаментал илмий тадқиқот лойиҳасида фойдаланилган (Ўзбекистон Республикаси Олий ва ўрта махсус таълим вазирлигининг 2019 йил 4 июлдаги 89-03-2685-сон маълумотномаси). Илмий натижалар асосида ўзбек луғатшунослиги ва миллий корпус лингвистикаси муносабати кўрсатиб берилган, луғавий-маъновий теглаш назариясидаги қарашлар ўрганилган ва умумлаштирилган;

ўзбек тили луғавий майдонлари асосида маъновий тег моделларини ишлаб чиқишнинг ҳамда ўзбек тилидаги омоним ва кўп маъноли сўзларни теглашнинг назарий асослари очиб берилган материаллардан 2012-2014 йилларда бажарилган ИТД-1-07 рақамли «Тилшунослик асослари фанидан дарслик, ўқув-услубий мажмуа ва уларнинг электрон нусхаларини яратиш» мавзусидаги амалий грант лойиҳасида фойдаланилган (Ўзбекистон Республикаси Олий ва ўрта махсус таълим вазирлигининг 2020 йил 19 июндаги 89-03-2152-сон маълумотномаси). Тадқиқот натижалари ўзбек тили корпуслари яратишнинг назарий асосларини ишлаб чиқишда хизмат қилган;

ўзбек тили атов бирликлари учун маъновий тег категориялари ҳамда константа ва оператор тег моделлари тузишнинг алгоритмини кўрсатиб берувчи материаллардан 2018-2019 йилларда Ўзбекистон миллий телерадиокомпанияси «Маҳалла» телеканалининг «Маърифат ёғдуси» ва «Машварат» кўрсатувларининг сценарийларини тузишда фойдаланилган (ЎзМТРКнинг 2020 йил 10 февралдаги 01-18-132-сон маълумотномаси). Натижада томошабинлар дунёқарашида ўзбек тили маъновий майдонлари асосида маъновий тег моделлари ишлаб чиқиш ҳақидаги тасаввурларнинг пайдо бўлиши, ёшларнинг билим олишга бўлган қизиқиши ҳамда интернет орқали ўзбек тилига оид билимларининг такомиллашишига кўмаклашилди.

Абдулла Қаҳҳорнинг «Анор» ҳикояси корпуси интерфейси лойиҳалаштирилган; асар матни атов бирликлари луғавий-маъновий теглари лингвистик моделидан корпус on-line варианты фрагментини яратишда фойдаланилган (Ўзбекистон Республикаси Адлия вазирлиги ҳузуридаги Интеллектуал мулк агентлигининг № DGU 08980 рақамли гувоҳномаси). Натижада «Абдулла Қаҳҳорнинг «Анор» ҳикояси семантик разметкаси» on-line варианты фрагменти яратилган.

**Тадқиқот натижаларининг апробацияси.** Мазкур тадқиқот натижалари 5 та халқаро, 7 та республика илмий-амалий анжуманларида апробациядан ўтган.

**Тадқиқот натижаларининг эълон қилинганлиги.** Диссертациянинг асосий мазмуни муаллиф томонидан чоп этилган жами 20 та илмий иш, жумладан, Ўзбекистон Республикаси Олий аттестация комиссиясининг докторлик диссертациялари асосий илмий натижаларини чоп этиш тавсия этилган илмий нашрларда 8 та мақола, жумладан, 5 та республика ҳамда 3 та хорижий журналларда нашр этилган.

**Диссертациянинг тузилиши ва ҳажми.** Диссертация кириш, уч боб, хулоса, фойдаланилган адабиётлар рўйхати ва иловадан иборат. Диссертациянинг умумий ҳажми 146 саҳифани ташкил этади.

## ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

**Кириш** қисмида мавзунинг долзарблиги асосланган, тадқиқотнинг мақсад ва вазифалари, объекти ва предмети тавсифланган, республика фан ва технологиялари ривожланишининг устувор йўналишларига мослиги кўрсатилган, илмий янгилиги ва амалий натижалари баён қилинган, олинган натижаларнинг илмий ва амалий аҳамияти очиб берилган, натижаларни амалиётга жорий қилиш, нашр этилган ишлар ҳамда диссертация тузилиши бўйича маълумотлар келтирилган.

Диссертациянинг «Ўзбек тили атов бирликлари орасидаги лексик-семантик муносабатлар» деб номланувчи биринчи бобида ўзбек тилида лексик-семантик майдон назариясининг корпус тузишдаги аҳамияти, лексик-семантик муносабатларнинг атов бирликларини теглашдаги ўрни ҳамда ўзбек лексикографияси ва корпус лингвистикаси муносабати масалалари ўрганилган. Бобнинг «Ўзбек тилида лексик-семантик майдон назариясининг корпус тузишдаги аҳамияти» номли биринчи бўлимида тилга систем ёндашув тарихи, ўзбек тилшунослигида синонимик қатор, лексик-семантик гуруҳ, лексик-семантик майдон ҳақидаги тадқиқот натижаларидан корпус семантик тегларини тузишда фойдаланиш, корпус материалларини теглашда ЛСГ, МГ ва ЛСМнинг аҳамияти масалалари кўриб чиқилган.

Тил – яхлит бир бутунликни ташкил қилувчи элементлар системаси. Уни ташкил этувчи ҳар бир элемент маъноси бир вақтнинг ўзида системани ташкил қилаётган бошқа элементларнинг мавжуд бўлишидан келиб чиқади<sup>10</sup>.

<sup>10</sup> Соссюр Ф. Де. Труды по языкознанию. – Москва: Прогресс, 1977. – С.147.

В.Гумбольдт, А.Шлейхер, Г.Штейнталь, В.Матезиус, В.Скаличка, Н.С.Трубецкой, Л.Ельмслев, Х.Ульдаль, Л.Блумфильд, Э.Сепир, Б.Уорф, З.С.Харрис, А.Мартине, Л.Вайсгербер, Бодуэн де Куртенэ, Ф.Фортунатов сингари жаҳон тилшунослигининг кўзга кўринган олимлари томонидан илгари сурилган лингвистик концепцияларда тилга систем-структур ёндашувнинг турли хил кўринишларини учратиш мумкин<sup>11</sup>. Рус тилшунослигида лексик система ва структура етарлича тадқиқ қилинган<sup>12</sup>, ўзбек тили лексикасини атрофлича ўрганишда ўзбек олимларининг хизматлари беқиёс. Улар фундаментал тадқиқотлари билан ўзбек тили лексикасини система сифатида ўрганиш учун улкан замин ҳозирладилар; бу усулни илк босқичда бошлаб ҳам бердилар; ўзбек тили систем лексикологиясига асос солдилар<sup>13</sup>, бир гуруҳ олимларнинг илмий изланишлари эса тўғридан-тўғри лексикани систем тилшунослик нуқтаи назаридан ўрганишга бағишланган<sup>14</sup>.

Луғавий семантик майдон (ЛСМ) назарияси асосида соҳалараро алоҳида илмий-тадқиқот ишлари олиб бориш, асосан, XX асрнинг 70-80-йилларида бошланди. Бу босқич лексикани маълум бир семантик гуруҳлар (ЛСГ)дан, муайян маъно элементларининг муносабатларидан ташкил топган яхлит система сифатида ўрганиш билан характерланади.

ЛСМ ўзида бир неча ЛСГларни қамраб олган нисбатан кенгроқ майдон. Сўзни синонимик қатор, ЛСГ ёки ЛСМга қамраб олишда семанинг умумий базасига, шунингдек, гуруҳ таркибидаги ҳар бир лексик birlikнинг

<sup>11</sup>Қаранг: Основные направления структурализма. – Москва: Наука, 1964. – 360 с.; Звегинцев В.А. История языкознания XIX и XX веков в очерках и извлечениях. – ч. I. – Москва, 1964.; ч. II. – Москва, 1965. – 496 с.; Звегинцев В.А. Язык и лингвистическая теория. – Москва: Издательство МГУ, 1973. – 248 с.; Философские основы зарубежных направлений в языкознании. – Москва: Наука, 1977. – 295 с.; Новое в лингвистике. – Москва: Прогресс. Вып. I. 1960.; Вып. 2. 1962.; Вып. 3. 1963. – 686 с.

<sup>12</sup>Қаранг: Щерба Л.В. Языковая система и речевая деятельность. – Ленинград, 1974. – 418 с.; Виноградов В.В. Избранные труды. Лексикология и лексикография. – Москва: Наука, 1977. – 360 с.; Ахманова О.С. Очерки по общей и русской лексикологии. – Москва: Учпедгиз, 1957. – 194 с.; Уфимцева А.А. Опыт изучения лексики как системы (на материале английского языка). – Москва: Издательство АН, 1962. – 288 с.; Степанов Ю.С. Семиотика. – Москва: Радуга, 1983. – 627 с.; Шмелев Д.Н. Проблемы семантического анализа лексики. – Москва, 1973. – 280 с.; Новиков Л.А. Семантика русского языка. – Москва: Высшая школа, 1982. – 272 с.

<sup>13</sup>Ибрагимов С. Профессиональная лексика узбекского языка: Автореф. дис... д-ра фил. наук. – Ташкент, 1961. – 41 б.; Ҳожиев А. Ўзбек тили синонимларининг изоҳли луғати. – Тошкент: Ўқитувчи, 1974. – 308 б.; Раҳматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. – Тошкент: Ўқитувчи, 1978. – 406 б.; Қўчқортюев И. Сўз маъноси ва унинг валентлиги. – Тошкент: Фан, 1977. – 187 б.; Бегматов Э. Ҳозирги ўзбек адабий тилининг лексик қатламлари. – Тошкент: Фан, 1985. – 199 б.; Пинхасов Я. Ҳозирги ўзбек адабий тили (Лексикология ва фразеология). – Тошкент: Фан, 1969. – 286 б.; Дадабоев Х. Военная лексика в староузбекском языке: Автореф. дис.... канд. фил. наук. – Тошкент, 1981.; Бафоев Б. Кўҳна сўзлар тарихи. – Тошкент: Фан, 1991. – 152 б.

<sup>14</sup>Раҳматуллаев Ш. Семема – мустақил тил бирлиги // Ўзбек тили ва адабиёти. – 1984, 5-сон.; Қўчқортюев И. Сўз маъноси ва унинг валентлиги. – Тошкент: Фан, 1977. – 187 б.; Бегматов Э., Неъматов Х., Расулов Р. Лексик микросистема ва унинг тадқиқ методикаси (Систем лексикология тезислари) // Ўзбек тили ва адабиёти. – 1989, 6-сон. – Б.35- 40.; Неъматов Х., Расулов Р. Ўзбек тили систем лексикологияси асослари. – Тошкент: Ўқитувчи, 1995. – 128 б.; Расулов Р. Ўзбек тилида ҳолат феъллари ва уларнинг облигатор валентликлари. – Тошкент: Фан, 1989. – 131 б.; Бозоров О. Ўзбек тилида даражаланиш: Фил. фанлари д-ри ... дис. автореф. – Тошкент, 1997.; Искандарова Ш. Ўзбек тили лексикасини мазмуний майдон сифатида ўрганиш (шахс микромайдони): Фил. фанлари д-ри ... дис. автореф. – Тошкент, 1999.; Сафарова Р.Г. Гипонимия в узбекском языке (на материале общеупотребительных зоонимов): Автореф. дис.... канд. фил. наук. – Тошкент, 1990.; Орифжонова Ш. Ўзбек тилида луғавий градуонимия: Фил. фан. номзоди ... дис. автореф. – Тошкент, 1996. – 58 б.; Қиличев Б. Ўзбек тилида партонимия: Фил. фан. номзоди ... дис. автореф. – Тошкент, 1997.

маъно(лар) тизимига асосланилади.

Сўзларнинг яна ҳам каттароқ доирадаги гуруҳлашуви – мавзувий гуруҳ (МГ); бу гуруҳ турли сўз туркумидаги сўзларнинг мавзу умумийлиги асосидаги тўплами. Бу тўплам турли: парадигматик ва синтагматик муносабатлар асосида тузилади. Масалан, «спорт» МГ (*футбол, гол урмоқ, стадион, мухлис*) ёки «савдо» МГ (*савдо қилмоқ, савдолашмоқ, бозор, магазин, сотувчи, харидор, сотув, сотмоқ*). МГга турли ЛСГлар киради. Масалан, *савдо биноти* ЛСГи (*магазин, киоск, бутик, супермаркет, дўкон*), синоним (*эга бўлмоқ, сотиб олмақ*), антоним (*қиммат – арзон*), гипоним (*магазин – гастроном*)лар «савдо» МГга киради.

Тил эгалари томонидан бу бирликлар нутқда аралаш ҳолда қўлланилаверади. Тил корпусларини семантик теглашда ЛСГ, МГ ҳамда ЛСМнинг чегараси аниқ, қидирув ойналарида ушбу иерархия тўғри жойлашган бўлиши лозим. «МГ олам лисоний манзарасини, унинг фрагментларини ифодаласа, ЛСМ маъно қирралари ва муносабатларини ифодалайди»<sup>15</sup> принципига амал қилишимиз ўринли.

Лексикани ўрганишга икки: семасиологик (сўздан тушунчага) ва ономасиологик (тушунчадан сўзга) аспектдан ёндашиш тил корпусида сўз билан ишлаш, сўз ўрганиш, лингвистик амаллар бажаришнинг асосий йўналишларини белгилаб беради. Сўзга мавзувий гуруҳ, лексик-семантик гуруҳ, семантик майдон асосида ёндашиш корпусда тушунчадан сўзга қараб бориш имкониятини беради. Жаҳондаги мавжуд семантик аннотация (разметка, тег)га эга корпус<sup>16</sup>ларнинг кўпчилигида «тушунчадан сўзга» тамойилига амал қилинган. Бизнингча, корпус материални семантик теглашда икки томонлама: сўздан тушунчага, тушунчадан сўзга қараб бориш ниҳоятда катта аҳамиятга эга. Корпусда бирликлар семантик жиҳатдан ЛСГ > МГ > ЛСМ иерархик тарзда қидирилади. Бу – қидирувнинг «тушунчадан сўзга» принципи. Қидирувнинг биринчи ойнасида ЛСМлар акс этади; ҳар бир ЛСМнинг ичига кирилганда яна кичик ЛСМлар жойлашади; уларнинг ҳар биттаси МГ, МГлар эса ЛСГлардан иборат бўлади. Корпус материални семантик теглаш ва қидирув ойнасида ЛСМларни жойлаштириш бир неча босқичда амалга оширилади.

Бобнинг «Лексик-семантик муносабатларнинг атов бирликларини теглашдаги ўрни» номли иккинчи бўлимида ўзбек тилида лексик-семантик муносабат турлари, уларнинг ўрганилиши; ўзбек тили бирликлари ўртасидаги лексик-семантик муносабат асосида семантик теглар мажмуини яратиш муаммолари таҳлилга тортилган.

Лексик бирликлар лисонда турли семантик муносабат асосида турлича парадигма ҳосил қилган ҳолда мавжуд бўлади. Синонимик, антонимик, градуонимик, партонимик, гипонимик муносабат ана шундай лисоний

<sup>15</sup><https://studfiles.net/preview/5837469/page:4/>: (Рублева О. Л.. Лексикология современного русского языка. – Москва, 2016. – 257 с.)

<sup>16</sup> Қаранг: (<http://ruscorpора.ru/corpора-sem.html>)

муносабат турларига киради<sup>17</sup>. Тил бирликларининг диалектика қонуниятлари асосида бир ерга жамланиши, алоҳида лексик-семантик гуруҳларни ташкил қилиши, гуруҳ ичидаги бошқа аъзолардан маълум бир белгилар асосида фарқланиши сабаблари кенг қўламда тадқиқ қилинган. Манбаларда семантик майдондаги парадигма аъзолари ўртасидаги муносабат турлари синонимик, гипер-гипонимик, холо-меронимик, градуонимик, функционимик, зидлик муносабати каби гуруҳларга бўлинган<sup>18</sup>.

Маълумки, тил корпуси – лингводидактик, лингвистик ва бошқа ижтимоий соҳаларнинг зарурий иш қуроли. Корпус ҳақидаги тадқиқотларда корпуснинг сўз ўрганишда катта ахборот базаси эканлиги аллақачон ўз исботини топган<sup>19</sup>.

Семантик теглашда «сўздан тушунчага» принципи сўз маънолари изоҳи, лексик-семантик муносабатга киришувчи лексемалари: синоним, антоним, гипоним/гипероним, градуоними, холо/мероними ҳақидаги изоҳларни тег сифатида бириктиришдан иборат. Семантик теглар тизимига сўзшакл, лемма, изоҳли луғатдаги маънолари, мавжуд синоним, градуоним, гипоним/гипероним, холо/мероним, антонимлари ҳақидаги изоҳ киради. Семантик тег таркибига *туркумий белгиси, маъноси, синоними, антоними, уяси, уядоши, даражаланиш қатори, келиб чиқиши, омоними, услуги, пароними, варианты, қўлланиш даври* каби белгилар кириши ўринли. Айтиш мумкинки, тил бирликларини корпус учун семантик теглашда шу пайтгача тилшуносликда лексик-семантик муносабатлар тадқиқи натижасида қўлга киритилган хулоса, назарий материал, луғатлар лингвистик таъминот базаси сифатида хизмат қилади.

Бобнинг «Ўзбек лексикографияси ва корпус лингвистикаси» деб аталган учинчи бўлимида луғат мақоласи ва семантик теглар мажмуидаги муштарак, фарқли жиҳатлар, ўзбек тилидаги луғатлардан корпус материали сифатида фойдаланиш аҳамияти, тил корпуси учун ахборот манбаи бўла оладиган луғатлар, луғат мақоласининг тузилиши, уларнинг корпус бирлигини теглашдаги имконияти очиб берилган.

Ўзбек тилида мавжуд луғатлар матнини электронлаштириш – луғатни корпус билан боғлашнинг илк босқичи. Лексикада системалилик тилнинг

---

<sup>17</sup>Sayfullayeva R. va b. Hozirgi o`zbek adabiy tili. O`quv qo`llanma. – Toshkent : Fan va texnologiya, 2009. – B.112.

<sup>18</sup>Собиров А. Ўзбек тилининг лексик сатҳини системалар системаси тамойили асосида тадқиқ этиш. – Тошкент: Маънавият, 2004. –140 б.

<sup>19</sup>Бу ҳақда тўлиқроқ маълумот олиш учун қаранг: Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – 48 с. – С.27.; Кутузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) – //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf; Национальный корпус русского языка // Вопросы языкознания. – Москва: Наука, 2006. – С.149-155.; [http://wikipedia.org/wiki/ Корпусная лингвистика](http://wikipedia.org/wiki/Корпусная_лингвистика). Статья в интернет-энциклопедии Википедия; [www.ruscorpora.ru/](http://www.ruscorpora.ru/) Плу́нган В. Зачем мы делаем корпусы? “Отечественные записки” 2005, №2. [http://magazines.russ.ru/oz/2005/2/2005\\_2\\_20-pr.html](http://magazines.russ.ru/oz/2005/2/2005_2_20-pr.html); Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Фил.фан.бўйича фалсафа доктори (PhD)...дис. – Бухоро, 2018. – 250 б.

бошқа сатҳларидаги каби яққол кўзга ташланиб турмайди. Лексикани сифат ва миқдор жиҳатдан инвентаризация қилиш, кенг тадқиқ қилиш имконияти пайдо бўлган. Тил бирликларини семантик теглаш учун энг асосий манба ўзбек тилининг изоҳли луғати<sup>20</sup> саналади. Луғатда бош сўзнинг ҳар қандай маъноси эмас, балки унинг ҳозирги умумистеъмолда бўлган, кўпчиликка тушунарли бўлган маъноси/лари қайд этилган. Бош сўз бир маъноли бўлса, луғат мақоласи шу маъносини изоҳлаган; кўп маъноли бўлса, ҳар бир маъноси, маълум тартибда, тўқ қора рангли араб рақами билан алоҳида-алоҳида қайд этилган. «Сўздан тушунчага» принципидаги теглашда сўз маъноларини белгилаш луғатдаги тартибда амалга оширилиши айни муддао. Аммо «тушунчадан сўзга» принципидаги қидирувда кўп маъноли сўзнинг маъносини тегишли ЛСГ, МГ ва ЛСМларга бириктириш лозим, чунки кўп маъноли сўз бир маъноси билан бир ЛСГга мансуб бўлса, иккинчи маъноси билан бошқа ЛСГ/ЛСМга тегишли бўлади.

Луғат мақоласида иллюстратив мисоллар миқдори чегараланган бўлади; корпусда аксинча: қидирилган бир сўзнинг минглаб мисоллар контекстига ҳавола қилиш имкони бор. Мавжуд антоним ва синонимлар луғатини корпуснинг базасига боғлаш, қидирув берилганда шу луғатлардан гиперсилка асосида топишни йўлга қўйиш яхши натижа беради. Гипергипонимик, холо-меронимик муносабатни акс эттирувчи луғат ўзбек тилида ҳали яратилмаганлиги сабабли сўзни ушбу параметрлар асосида семантик теглаш учун лексикографик материал мавжуд эмас. Семантик теглашда бу борада қилинган тадқиқотларга таянган ҳолда мустақил, қўл меҳнати билан бажарилади.

Тадқиқотнинг иккинчи боби «**Атов бирликларини лексик-семантик теглашнинг назарий асос ва моделлари**» деб номланган. Бобнинг «Лексик-семантик теглаш назариясидаги қарашлар» сарлавҳали биринчи бўлимида семантик разметка ҳақида умумий маълумот келтирилган, семантик разметканинг ўрганилиши, амалиётда қўлланиши таҳлил этилган.

Семантик тег – тил корпусидаги сўз ёки сўз бирикмасининг маълум бир семантик категория ёки ундан ҳам кичикроқ семантик гуруҳ (ЛСГ, семантик майдон ва тўда)га мансублигини билдирувчи, маънони махсуслаштирувчи белги, изоҳлар мажмуи<sup>21</sup>. Корпуснинг семантик теглари сўз маъноси(лари) спецификаси, сўзнинг омонимлик, синонимлик билан боғлиқ изоҳлар мажмуини тузиш, сўзни категориялаш, унинг мавзувий гуруҳ, ЛСГ, семантик майдонга мансублигини белгилаш, деривацион характеристикаси, аташ маъносини қамраб олади. Мавжуд миллий корпуслар семантик теглаш принципларини кузатиш корпусда бир сўзшаклга уч турдаги белги (разряд, лексик-семантик характеристика, деривацион тавсиф) бириктирилишини кўрсатди. Лексик-семантик теглар (1) таксономия (лексема мансуб ЛСГ) – от,

<sup>20</sup> Ўзбек тилининг изоҳли луғати (А.Мадвалиев таҳрири остида). 5 жилдди. – Тошкент: “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти, 2000-2006.

<sup>21</sup> Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с. – С.48-49.

сифат, феъл, равиш сўз туркумлари учун тегишли тег; (2) мереология («бутун-бўлак»ка ишора қилувчи белги, «элемент-тўда/гуруҳ»га мансублик белгиси) – предмет ва нопредмет атов бирликларига тегишли тег; (3) топология (ифодаланаётган объектнинг топологик мавқеи) – нарса отларига тегишли тег; (4) каузация – феълларга тегишли тег; (5) баҳо – предмет ва нопредмет атов бирликлари, сифат ва равишга тегишли тег майдони бўйича гуруҳланади.

Семантик теглашда ягона стандарт шакл бўлмаса ҳам, ҳарф, рақам ёки фақат рақамдан иборат кодлардан фойдаланилади. Биринчи ҳарф ёки рақам умумий семантик маънони, кейинги белги сўз маъносини янада махсуслаштирувчи кичик семантик гуруҳни ифодалайди. Семантик тег нафақат сўз, балки кўплаб бирикмаларни семантик гуруҳга бирлаштиради, бундай пайтда турли бирикувдаги бир маънони билдирувчи бирикмалар битта белги билан кодланади. Семантик теглаш корпус воситасида лингвистик амаллар бажариш имконини кенгайтирувчи иш қуроли: корпусда семантик аннотацияланган бирликнинг мавжуд бўлиши фақат алоҳида сўзни эмас, балки фойдаланувчи томонидан сўралган бирикма ҳолидаги конструкция кидирувига жавоб бериш имконини пайдо қилади. Семантик разметка морфологик разметканинг давоми ва кенгайтмаси саналади; у ўзида уч гуруҳ белгилар тизимини: (1) сўз ясовчи кўрсаткичлар билан шаклланган бирлик; (2) лексик-семантик хусусиятга тегишли белги: атоқли от, нарса оти, аслий ва нисбий сифат; ўз ва кўчма маъно; (3) хусусий-семантик белги: мавзувий (таксономик) гуруҳ, семантик майдонни акс эттиради<sup>22</sup>. Корпусни семантик теглаш дастурчи ва тилшунос компетенцияси бирлашиши асосида икки асосий кўринишга ажратилади:

- 1) терминологик разметка – матндаги тушунчани номлаш босқичи;
- 2) бирликлараро муносабатни кўрсатувчи тег билан таъминлаш. Бунда ҳар қайси бирликнинг семантик майдондаги ўрни, бошқа бирлик билан муносабатини аниқлаш назарда тутилади.

А.А.Кретов тегларни константа, оператор-классификаторга ажратади<sup>23</sup>. Масалан, *Sp t:constr* бино ва иниоот (уй, чордоқ, кўприк). Ушбу теглар тизимида *Sp* константа, *t:constr* эса классификатор.

Бобнинг «Корпус лингвистикасида семантик тег категориялари: тажриба ва таҳлил» номли бўлимида семантик тег тизимида турли ёндашув, қарашлар; семантик разметка тизимида тег гуруҳлари тадқиқ этилган. Корпус лингвистикасида семантик разметка, унинг теглар тизими, тег категориялари, семантик теглаш муаммолари, кўп маънолилик ва омонимлик муаммосини ечиш масаласига оид бир қатор тадқиқотлар вужудга келган<sup>24</sup>. В.В.Куканова

<sup>22</sup>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.

<sup>23</sup>Кретов А.А.Анализ семантических помет в НКРЯ // <http://ruscorpora.ru/sbornik2008/11.pdf>.

<sup>24</sup>Куканова В. В. Принципы семантической разметки национального корпуса калмыцкого языка // [http://kalmcorp.ru/sites/default/files/kukanova\\_25.pdf](http://kalmcorp.ru/sites/default/files/kukanova_25.pdf); Кретов А.А. Анализ семантических помет в НКРЯ // <http://ruscorpora.ru/sbornik2008/11.pdf>; 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:

«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 маъносини аниқ, тўлиқ тавсифлашни, унинг қўлланиш контекстини кўрсатади. Иккинчи ёндашувда эса, аксинча, айти бир мазмунни ифодадай оладиган барча шакллари бир нуқтага йиғиш мақсад қилинади: бунда қайсидир маъно мажмуи тўпланади. Иккала ёндашув ҳам тилшунослик ва тил таълими учун бирдек аҳамиятли. Грамматика ва луғат ўқувчиси она тили (хорижий тил)ни 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 демокчилигини билади, ҳис қилади, аммо керакли сўз/грамматик шаклни топа олмайди»<sup>25</sup>. Ҳақиқатан, семантик разметка тил лексикаси тадқиқи учун жуда муҳим: тадқиқотчи/фойдаланувчи/ўқувчига сўзнинг бириктиш имконияти, синтактик конструкциясини кузатишга кенг йўл очади. Масалан, айрим сифатлар борки, фақат маълум отлар билан семантик валентлик имкониятига эга. Интерфейсда шундай сўрловларни ҳосил қилиш мисоллар массиви орқали сўзнинг валентлигини кўрсатади.

Масалан, *шайтон* сўзи контекстга қараб қуйидаги маъноларни билдиради:

ШАЙТОН I [иблис; жин, ажина] 1 дин. Одамларни дин йўлидан оздирувчи, уларни гуноҳга, жиноятга, разолатга бошловчи ёвуз рух ёки ёвуз рухларнинг бошлиғи; иблис. 2 кўчма Айёр, қув одам; алдамчи. *Анави магазинчи бор-ку. Ўлгудек шайтон.* Р.Файзий, Чўлга баҳор келди. 3кўчма Шўх, ўйноқи. ...*Ноумид – шайтон... Шайтон арава ... Шайтон йиғиси ... Шайтонга ҳай бермоқ ... Шайтони зўр (ёки кучли) ... Шайтони қўзимоқ*<sup>26</sup>.

Бу жараённинг алгоритми, тахминан, қуйидагича бўлади:

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

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

---

Индрик, 2005. – С.155-174.; 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.). Санкт-Петербург: Санкт-Петербургский гос.университет, Факультет филологии и искусств, 2008. – С.276-284.; 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». – Москва, 2006. – С.445-450.; 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. Плуноян. Санкт-Петербург: Нестор-История, 2009. – С.215-239.;

<sup>25</sup>Куканова В.В. Принципы семантической разметки национального корпуса калмыцкого языка // [http://kalmcorp.ru/sites/default/files/kukanova\\_25.pdf](http://kalmcorp.ru/sites/default/files/kukanova_25.pdf)

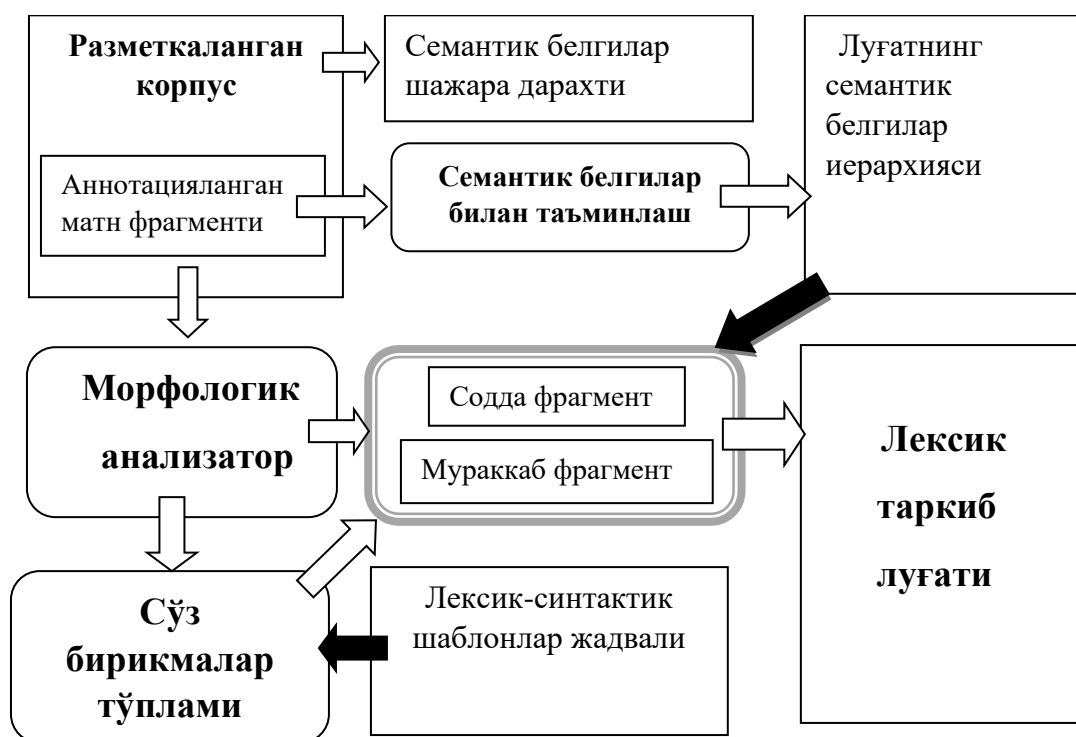
<sup>26</sup>Ўзбек тилининг изоҳли луғати (А.Мадвалиев таҳрири остида). 5 жилдли. – Тошкент: “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти, 2000-2006. – 4-том. – 534-б.



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, жинойтга, разолатга бошловчи ёвуз руҳ ёки ёвуз руҳларнинг 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* бирор хатти-ҳаракатдан ўзини тиймоқ, ўзини босмоқ; *шайтони зўр* шўхлиги тез ва кучли кўзғаладиган, шундай хусусиятли; *шайтони қўзимоқ* шайтони кўзгамоқ, жини тутмоқ.

Семантик разметканинг ўзига хос томони аннотациялашнинг икки хил: фасет ва шажара усулининг мавжудлигида. Амалда бирликнинг хусусиятига қараб 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, интерфейсда акс этадиган қидирувда ҳам сўз → маъно, маъно → сўз принципига амал қилишини кўрсатади. Разметка воситаларининг таркибий қисмлари ва уларнинг ўзаро муносабатини қуйидаги жадвалда кузатиш мумкин:



Бобнинг «Семантик разметка тизими ва моделларини тузиш алгоритми» деб номланган учинчи бўлимида семантик разметка тизими, модели, семантик разметка тизимида кўп маънолилик ва филтър масалалари муҳокама этилган.

Кўп маънолиликни фарқлаш учун катта контекст керак: сўзловчи ва тингловчи бундай кўп маънолиликни нутқда сезмайди. Корпус разметкаси, одатда, кўп маънолиликни аниқлашда бирликнинг «кўшни» сўзга бирикишига қараб иш тутади. Бундай алгоритмни яратиш учун қоидалар мажмуи (типик ва кам учрайдиган ҳолатларнинг барчаси) тилшунос томонидан лингвистик таъминот базасига киритилиши лозим. Разметка тизимида бирлик разрядини белгилаш сўзнинг нарса/шахс оти билан бирика олиш қобилиятига қараб фарқланади: *енгил* сифати нарса оти билан бирикма ҳосил қилганда «оғир эмас» маъносини англатади. Бу бирликнинг кейинги разрядлари эса нопроект отлар таксономлари билан чегараланади. Унинг филтърларидан бири қуйидаги кўринишга эга бўлади:

Сўз	Контекст	Яқуний маъно
<i>Енгил</i>	+ нарса оти	Sem= разряд «белги-хусусият» таксономик синф: физик хусусият: вазн

*Яланғоч* сўзининг иккита маъноси мавжуд. Қиёсланг: *Яланғоч 1* – яланғоч киши; *Яланғоч 2* – яланғоч фикр.

Бирликнинг ҳар бир маъноси отнинг таксономик синф билан бирикишида чегараланади. Семантик разметкаси қуйидаги шаклга эга бўлиши мумкин:

Сўз	Контекст	Яқуний маъно
<i>Яланғоч</i>	+ шахс	Sem= разряд «белги-хусусият» таксономик синф: физик ҳолат
<i>Яланғоч</i>	+ ўрин-жой	Sem = разряд «белги-хусусият» таксономик синф: ташқи кўриниш
<i>Яланғоч</i>	+ нарса оти	Sem= разряд «белги-хусусият» таксономик синф: физик хусусият

Фикримизни далиллаш учун *совуқ* сўзини таҳлил қиламиз. Бу бирликнинг *совуқ шамол* (паст температура), *совуқ ранг* (кўримсиз ранг), *совуқ нигоҳ* (инсон психологик ҳолати) каби маънолари мавжуд. Семантик валентлигини кўйидаги контекстда кузатиш мумкин:

Сўз	Контекст	Яқуний маъно
<i>Совуқ</i>	+ табиат ҳодисаси + вақт	Sem= разряд «белги-хусусият» таксономик синф: физик хусусият: температура
<i>Совуқ</i>	+ ранг	Sem= разряд «белги-хусусият» таксономик синф: физик хусусият: ранг-тус
<i>Совуқ</i>	+ ментал соҳа + психик соҳа + инсонга хос хусусият + инсоннинг ҳаракати ва хулқи	Sem= разряд «белги-хусусият» таксономик синф: инсонга хос хусусият

Корпус семантик разметкаси учун зарурий воситалар сифатида ўзида маълум бир тилнинг лексикасини акс эттирувчи сўзлик (1), тил лексикасини тўла изоҳлай оладиган семантик луғат (2), семантик разметкани амалга ошириш учун лингвистик модул – қоидалар мажмуи (3), семантик разметка тизими (4), кўшимча дастурий восита: кўп маънолилиқ ва омонимликни фарқлай оладиган филтер (5) талаб қилинади. Семантик теглаш жараёнида морфологик, лексик омонимлик; кўшма сўз (таркибли термин) таркибига кирувчи универсал лексика; луғатларда мавжуд бўлмаган сўз; бўлак фрагмент; ҳарфий-рамзий конструкцияларни теглашнинг ўзига хос тамойилларини ишлаб чиқиш лозим, чунки ушбу бирликлар ҳар бир тилда алоҳида характерга эга бўлади.

Диссертациянинг учинчи боби «**Ўзбек тили корпуслари лексик-семантик теглари лингвистик таъминотини яратиш муаммолари**» деб номланган. Бобнинг «Тил бирликларини семантик теглашда семеманинг аҳамияти» номли биринчи бўлимида тил бирликларини семантик теглашда семема масаласи, ўзбек тили бирликларини семантик теглашда луғавий микросистема асосида ёндашув муаммоси таҳлилга тортилган.

Ҳар бир тилнинг корпуси учун семантик теглар мажмуини яратиш тамойилларини ишлаб чиқиш мазкур тилга хос хусусиятлардан келиб чиққанлиги сабабли ўзбек тили семантикаси, семасиологияси борасидаги тадқиқотлар ўзбек тили корпусларида семантик теглар мажмуини яратиш, сўзларни семантик теглаш тамойилларини ишлаб чиқиш учун назарий асос бўлади. Г.Глисон тилнинг уч компоненти: ифода структураси, мазмун структураси ва луғатни<sup>18</sup> ажратади. Ифода структурасининг бирлиги морфема бўлса, мазмун структурасининг бирлиги – семема. Г.Глисоннинг ушбу таснифи леммани семантик теглашда корпус лингвистикаси учун қўл келади. Тил корпусида сўзшаклни морфологик теглаш учун унинг ифода структурасига асосланган, семантик теглашда унинг мазмун структураси тўла тегланиши лозим. Агар мазмун структураси тўла қамраб олинмаса, тилимиз гўзаллигини тўлақонли намоён этиш имконияти бўлмайди.

Майдон чегарасини белгилаш корпус бирликларини семантик теглашда

хам аҳамиятга эга. Кўп маъноли сўзларнинг фақат бир майдонга мансублиги ёки бир неча майдоннинг элементлари бўлиши мумкинлиги сўзшакл ёки леммани теглашда бир қанча муаммоларни ечишни талаб қилади. Кўп маъноли сўзнинг қайси майдонга тегишли эканлигини семантик филтёр асосида аниқлаш ҳам мумкин бўлади.

Бобнинг «Ўзбек тилидаги полисемантик сўзларни теглашнинг назарий асослари» деб номланган иккинчи қисмида ўзбек тилида полисемия ҳодисасига ёндашув, семаларнинг лексикографик талқини ҳамда уларнинг корпусда берилиши масаласи тадқиқ этилган.

Корпус лингвистикасида кўп маъноли бирликни семантик теглаш жараёнида мунозарали қарашлардан холи, умумий принциплар ишлаб чиқилиши лозим. Корпус қидируви натижасида сўзнинг барча семалари кетма-кетлик асосида юзага чиқмайди, чунки қидирув натижаси белгилар кетма-кетлиги асосида амалга ошади. Агар тил корпусида семантик изоҳ маълумотлар омбори мавжуд бўлмаса, натижа фақат шакл асосида юзага чиқади. Семантик теглаш натижаси шундай бўлиши керакки, семалар кетма-кетликда кўринсин. Бунга фақатгина тил корпуси лингвистик таъминоти базасига ўзбек тилининг бирикувчанлик луғатини киритиш билангина эришиш мумкин. Бундай луғатнинг ўзбек тилида мавжуд эмаслиги ўзбек лексикографиясида яратилиши кечиктирилмаслиги лозим бўлган луғатлар қаторини яна биттага кўпайтиради. Зеро, ўзбек тили корпусларининг яратилиши ўзбек лексикографияси тараққиёти билан тўғри пропорционал. Тил корпусида тил бирликларининг нутқий воқеланиши акс этади: сўз тилда қандай қўлланилиши кузатилади. Қидирув натижаси мисол, контекст; аммо семалар кўрсатилмайди. Бизнингча, кўп маъноли сўзни семантик теглашда унинг умумистеъмол ва терминологик маъносини фарқлаш лозим. Қидирув натижасида интерфейсда аввал сўзнинг умумистеъмол маъноси, кейин терминологик маъноси кўриниши лозим. Бунда бирикувчанлик луғати асосида умумлашма тарзда тузиб чиқиладиган формула ва моделлар қўл келади.

Тил корпусида *одам* леммаси сўровда қандай натижа беради? Ушбу лемма ёки сўзшакл икки томондан кузатилиши мумкин. Агар сўров мақсади айнан шу «сўзни қидириш» бўлса, асос қидируви ёки сўзшакл қидирувини амалга ошириш мумкин. Тил корпуси базада мавжуд юзлаб *одам* сўзшакли учраган контекстларни топади ва унга ҳавола қилади. Фойдаланувчи *одам* лексемасининг маъноларини ажрата олмайди, фақат сўзнинг қўлланиш контекстлари, валентлик қобилияти ҳамда ишлатилиш частотасини аниқлаши мумкин. *Одам* лексемасининг ЎТИЛда изоҳланган 5та маъносини билдирувчи мисол ёхуд бу сўзнинг бир неча контекстда полисемантик табиатини намоён этишини кўрсатиш алгоритминини тузиш талаб этилади.

**Одам 1.** Фикрлаш, сўзлаш ва меҳнат қилиш қобилиятига эга бўлган, шу туфайли ҳамма махлуқотдан устун турувчи жонли зот, киши, инсон. *Ибтидоий одам.*

2. Кишилиқ жамиятининг айрим бир аъзоси, шахс. *Яхши одам. Ҳалол одам. Асл одам. Аблаҳ одам.*

3. Жамоа хўжалиги, оила аъзоларининг ҳар бири; жон.
4. Бирор бой-бадавлат ёки амалдор кишининг хизматкори ёки тарафдори.
5. Ўзга жамоа ёки оила вакили.

Юқоридаги изоҳларни ЎТИЛдаги тартибда келтирдик, тил корпусида (агар семантик тег мукаммал ишланмаган бўлса) шундай – полисемантик жиҳати фарқланмаган ҳолатда учрайди ва мисоллардан сўзнинг семаларини ажратиб олиш қийин. Луғат билан ишламаган фойдаланувчи бу ишни эплай олмайди. Айти бир матнда ушбу сўзнинг қайси маъноси воқелашганлиги фойдаланувчи учун муҳимми?

1. Фойдаланувчи сўзнинг қўлланиш контекстини, қайси сўзлар билан кўпроқ бирикишини аниқламоқчи бўлиши мумкин.

2. Фойдаланувчи сўзнинг семаларини, полисемантик табиатини кузатмоқчи бўлиши мумкин.

Биринчи ҳолатда тил корпусида қидирув натижасида чиқадиган контекстлар кетма-кетлигининг ўзи фойдаланувчининг муаммосини ҳал қилади. Иккинчи ҳолатда корпус тузувчи мутахассис томонидан бундай сўзлар махсус ишлов берилган бўлиши лозим. Тажрибалар кўрсатадики, полисемантик сўзларга ЎТИЛда берилган изоҳлар гиперссилка орқали бириктирилиши мумкин. Агар сўз полисемантик табиатли бўлса, қидирув натижаси алоҳида ранг билан ажралиб туриши лозим. Шундай сўзлар танланганда гиперссилкадаги ҳавола очилиб, фойдаланувчи бундай сўзларнинг семаларини, уларнинг нутқий воқеланишини худди ЎТИЛдагидек ўқиб олиши мумкин. Ҳозирча, корпус лингвистикасида сўзнинг маъно кирраларини фарқлашнинг бошқа ечими топилмаган.

Бобнинг «Ўзбек тилидаги омонимларни теглаш муаммолари» номли учинчи бўлимида омоним бирликларни теглаш, ўзбек тили бирликлари учун семантик теглар мажмуи: константа ва оператор/классификатор тегларни ишлаб чиқиш масалалари ёритилган.

Омолексемаларни филтрлаш (снятие омонимии) алгоритми диққатга сазовор жараён. *Боглар I – боглар II, богим I – богим II, богини I – богини II* каби шакллар ҳар қандай грамматик шаклни олганда ҳам омонимлигини сақлаб қолади, уни грамматик шакллар асосидаги филтр билан фарқлаш имкони бўлмайди: ўрин-жой оти қидирилганда ҳам, ўлчов бирлиги ёки тўдани билдирувчи сўз қидирилганда ҳам аралаш ҳолда мисоллар массивида интерфейсда кўринади, уни фойдаланувчи фақат контекстан ажратиб олиши мумкин.

Турли туркумга хос бўлган омонимларни эса филтр орқали ажратиш имкони мавжуд. Масалан, *дайди* сўзи сифат ёки феъл туркумига тегишли бўлади. Шундай ҳолатлар учун қуйидагича филтр тузиш мумкин. У нутқда *дайдиб юрмоқ, дайдимайдиган* шаклида воқеланар экан, эҳтимолий валентлик асосида филтрни қуйидагича белгилаш мумкин:

- 1) дайди + б/*майдиган* = феъл;
- 2) дайди + исталган сўз = сифат.

*енг* сўзи ҳам шундай филтр асосида фарқланиши мумкин:

енг I от. Кийимнинг қўлни қолаб турувчи қисми. *Енгли, енгсиз, енглик, енгча.*

енг II феъл. Голиб келмоқ, зўр чиқмоқ, устун келмоқ. *Сени иш енгмасин, сен ишни енг*<sup>27</sup>.

1) енг+ли/сиз/лик/ча=от;

2) енг+ма/гин/ди/май/гунча/гани/гач/ган= феъл.

Бундай сўзлар луғатлардан йиғилиб, семантик филтър базаси яратилиши лозим.

Ҳатто икки туркумга мансуб (масалан, от ва сифат) бўлган кўплаб омонимларни филтърлаш ҳам муаммоли бўлиши мумкин. От ва феъл туркумида ўзаро омонимлик бўлса, юқорида таклиф қилинган филтър натижа бериши мумкин, аммо от ва сифат туркуми орасида омонимлик бўлса, юқорида таклиф қилинган филтър натижа беролмайди. Масалан,

ёпиқ I сифат. Ёпилган ҳолатда, бекик, берк. *Эшик, деразалар ёпиқ.*

ёпиқ II от. От устига ёпиладиган мато, жул [попона]. *Сайислар ... отлар устидаги ёпиқларни ва тўқимларни олиб, ҳалиги тол дарахти остига ёзганлар*<sup>28</sup>.

Ёпиқ от бўлиб келганда, эгалик, келишик, кўплик кўшимчаларини олади, сифат бўлганда, фақат -роқ (қиёсий даража) кўшимчасини олиши мумкин. Бундай ҳолатни филтър фарқлаши мумкин. Аммо ёпиқ сўзи от бўлиб, бош келишикда келса, оддий/орттирма даражадаги сифат бўлганда филтър иш бермай қолади.

1) ёпиқ+лар/ни/нинг/га(қа)/да/дан = от;

2) ёпиқ+роқ = сифат; аммо

3) ёпиқ+ҳар қандай сўз = ?

Демак, матнда омонимликни фарқлаш учун энг зарурий восита – морфо-семантик филтър. Омонимни фарқлайдиган махсус дастур – морфо-семантик филтър яратиш тил корпусида омонимлар қидирувини тартибга солади. Бундай филтърни яратишда мавжуд омонимлар луғати ахборот банки вазифасини ўтайди. Луғатда мавжуд бир неча туркумга мансуб омонимлар ўрганилиб, ҳар бир шакл учун индивидуал-умумий филтър схемасини ишлаб чиқиш талаб қилинади. Бир туркумга мансуб омонимларни филтърлаш кейинги тадқиқотларда ечилиши зарур.

## ХУЛОСА

1. Лексикадаги «сўздан тушунчага» ҳамда «тушунчадан сўзга» ёндашуви тил корпусида сўз устида лингвистик амал бажаришнинг асосий йўналишларини белгилаб беради ва корпусда лемма ЛСГ, ЛСМ асосида тизимлаштирилади. Ваҳоланки, сўзни синонимик қатор, ЛСГ ёки ЛСМга камраб олишда семанинг умумий базасига, гуруҳ таркибидаги ҳар бир лексик

<sup>27</sup>Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. – Тошкент: Ўқитувчи, 1984. – 215 б.

<sup>28</sup>Ўша манба, 61-б.

бирликнинг маъно(лар) тизимига асосланиладики, бунда тилнинг бутун лексик таркиби ҳар бир тил учун оламнинг мукамал, махсус манзарасини акс эттирувчи, алоқада турувчи семантик майдонларга бирлашади. Шунингдек, атов бирликларини семантик теглаш учун лингвистик таъминот талаб қилинади, чунки семантик теглар мажмуи сўзшакл, лемма, микроЛСГ, макроЛСГга мансубликни ифодаловчи тег, МГ, ЛСМга тегишлиликни билдирувчи тегдан иборат бўлади.

2. Тил корпуслари сўзнинг барча маъноси, лексик-семантик муносабатга киришадиган сўзлари билан ёнма-ён изоҳланганлиги сабабли изоҳли луғат, тезаурусдан фарқ қиладики, бунда изоҳли луғатда фақат сўз маъноси изоҳланади; тезаурус сўзнинг нисбатан кўпроқ қиррасини очишга қаратилган бўлса-да, қидирувнинг соддалиги, имконияти чекланганлиги билан корпусдан фарқ қилади. Корпуслар сўзнинг қўлланиш контекстига ҳавола қилиши, сўзнинг бошқа бирлик билан боғлана олиш имкониятини намоён қилиши билан семантик белгиларнинг барчасини кўрсатмайди.

3. Тил корпусида (луғатдаги каби) бошқа сўзга ҳавола қилишга эҳтиёж бўлмайди, чунки сунъий интеллект катта ҳажмли хотирага эгаки, унда қоғоз исроф қилинмайди, ҳамиша ҳажмни исталганча кенгайтириш имконияти мавжуд бўлиб, ҳажмнинг катталиги фойдаланувчига ҳеч қандай қийинчилик туғдирмайди ва миллионлаб сўз атиги бир неча мегабайтда жойлашади. Луғат мақоласида иллюстратив мисоллар миқдори чегараланган бўлса, корпусда қидирилган бир сўзнинг минглаб мисоллар контекстига ҳавола қилиш имкони бўлади.

4. Тил корпусидаги сўз/бирикманинг маълум бир семантик категорияга мансублигини билдирувчи, маънони махсулаштирувчи белги, изоҳлар мажмуи семантик разметка бўлиб, корпуснинг семантик теглари сўз маъноси(лари) спецификаси, сўзнинг омонимлик, синонимлик билан боғлиқ изоҳлар мажмуини тузиш, сўзни категориялаш, унинг мавзувий гуруҳ, ЛСГ, семантик майдонга мансублигини белгилаш, деривацион характеристикаси, аташ маъноси каби белгиларни қамраб олади. Семантик разметка тизими разряд, лексик-семантик характеристика ҳамда деривацион тавсифдан иборатлиги учун лексик-семантик теглар таксономия, мереология, топология, каузация, баҳо майдонлари бўйича гуруҳланади.

5. Семантик разметканинг ўзига хос томони шундаки, аннотациялашнинг икки хил: фасет ва шажара усулидан фойдаланилади. Бирликнинг хусусиятига қараб ушбу икки усулдан бири ёки иккаласи аралаш ҳолда қўлланилади. Фасет усули бирликка кетма-кет изоҳ беришни тақозо этса, шажара усули бирлик мансуб бўлган гуруҳ, тўда, майдонни белгилашни талаб қилади.

6. Корпус семантик разметкаси учун зарурий восита сифатида ўзида маълум бир тилнинг лексикасини акс эттирувчи бирликлар: 1) сўзлик, 2) тил лексикасини тўла изоҳлай оладиган семантик луғат, 3) семантик разметкани амалга ошириш учун лингвистик модул – қоидалар мажмуи, 4) семантик разметка тизими, 5) қўшимча дастурий восита: кўп маънолилиқ ва омонимликни фарқлай оладиган филтёр ажратилади. Семантик теглаш

жараёнида морфологик, лексик омонимия; қўшма сўз (таркибли термин) таркибига кирувчи универсал лексика; луғатда мавжуд бўлмаган сўз; бўлак фрагмент; ҳарфий-рамзий конструкцияни разметкаланиши ўзига хос тамойилларини ишлаб чиқиш лозим, чунки ушбу бирликлар ҳар бир тилда алоҳида характерга эга бўлади.

7. Корпус бирликлари семантик майдон, тўда, гуруҳга мансублигига кўра оператор ва константа тег билан белгиланади. Зеро, майдон чегарасини белгилаш семантик теглашда аҳамиятга эга. Кўп маъноли сўзнинг фақат бир/бир неча майдоннинг элементи бўлиши мумкинлиги сўзшакл/леммани теглашда муайян муаммоларни ечишни талаб қилади, чунки кўп маъноли сўзнинг қайси майдонга тегишлилиги семантик филтер асосидагина аниқланади.

8. Тил корпуси лексик бирликнинг барча маъносини кўрсатиш имкони мавжудлиги билан қоғоз луғатдан фарқланади, бироқ қоғоз луғатлар тил корпуси учун асосий хом ашё ҳисобланади. Семантик теглаш орқали семаларнинг кетма-кетликдаги кўринишига эришиш лозим. Бунда тил корпуси лингвистик таъминотига ўзбек тилининг бирикувчанлик луғатини киритиш билан эришиш мумкин. Бундай луғатнинг ўзбек тилида мавжуд эмаслиги ўзбек лексикографиясида яратилиши зарур саналган луғат турини яна биттага кўпайтиради.

9. Жаҳон корпус лингвистикасида омоним бирликларни теглаш, матнни автоматик ўқиш жараёнида омонимияни бартараф этишни – 1) грамматик меъёрга ҳамда 2) статистикага асосланган усуллари фарқланади. Омонимия адабий нутққа ва сўзлашув нутқида нисбатан алоҳида-алоҳида белгиланиши керак, чунки луғатда, асосан, ҳозирги ўзбек тилига мансуб лексемалар тасвирланган, сўзлашув нутқида хос сўзлар чекланган миқдорда киритилган. Аммо тил корпусида адабий нутқ ва сўзлашув нутқида қўлланувчи омонимлар бир хилда қўлланадики, луғатда акс этмаган омоним шакллар тил корпусида кузатилиши мумкин. Матнда омонимликни фарқлаш учун энг зарурий восита – морфо-семантик филтер, уни яратиш тил корпусида омонимлар қидирувини тартибга солади.

10. Турғун тегларга семантик майдон номи сифатида қараш мумкин: иккинчи даражали теглар оператор/классификатор бўлиб, ЛСГларга тегишлиликни билдиради ҳамда сўз маъносини аниқлаштиришга хизмат қилади; константа тег лексик-семантик майдонга мансубликни ифодалайди. Тажрибалар шуни кўрсатадики, ажратилган ЛСМ, ЛСГлар константа ва константага мансуб оператор/классификатор тег бўла олади, ваҳоланки, кейинги тадқиқотларда ушбу теглар сираси бойитилиши, константа ва оператор/классификатор теглар сираси мукамаллаштирилиши табиий.



**УЧЕНЫЙ СОВЕТ PhD.03/27.02.2020.Fil.72.09  
ПО ПРИСУЖДЕНИЮ УЧЕНЫХ СТЕПЕНЕЙ ПРИ  
БУХАРСКОМ ГОСУДАРСТВЕННОМ УНИВЕРСИТЕТЕ**

---

**БУХАРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**АХМЕДОВА ДИЛЬДОРА БАХОДИРОВНА**

**ЛИНГВИСТИЧЕСКИЕ ОСНОВЫ И МОДЕЛИ ЛЕКСИКО-  
СЕМАНТИЧЕСКОЙ РАЗМЕТКИ НОМИНАТИВНЫХ ЕДИНИЦ  
ДЛЯ КОРПУСОВ УЗБЕКСКОГО ЯЗЫКА**

**10.00.01 – Узбекский язык**

**АВТОРЕФЕРАТ**  
**диссертации доктора философии (PhD) по филологическим наукам**

**Бухара - 2020**

Тема диссертации доктора философии (PhD) зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за № В 2019.2.PhD/FII 777

Диссертация выполнена в Бухарском государственном университете.

Автореферат диссертации на трёх языках (узбекском, русском, английском (резюме)) размещен на веб-странице Ученого совета ([www.buxdu.uz](http://www.buxdu.uz)) и на информационно-образовательном портале «ZiyoNet» ([www.ziyo.net](http://www.ziyo.net)).

**Научный руководитель:** Менглиев Бахтиёр Ражабович,  
доктор филологических наук, профессор

**Официальные оппоненты:** Раупова Лайло Рахимовна,  
доктор филологических наук, профессор  
Абжалова Манзура Абдурашатовна,  
доктор философии по филологическим наукам

**Ведущая организация:** Каршинский государственный университет

Защита диссертации состоится 23 октября 2020 года в 10<sup>00</sup> часов на заседании Ученого совета PhD.03/27.02.2020.FII.72.09 по присуждению учёных степеней при Бухарском государственном университете. (Адрес: 200118, город Бухара, улица М. Икбола, 11. Тел.: (0 365) 221-29-14; факс: (0 365) 221-29-14, email: [buxdu\\_rektor@buxdu.uz](mailto:buxdu_rektor@buxdu.uz)).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Бухарского государственного университета (зарегистрирована за № 968). Адрес: 200118, город Бухара, улица М. Икбола, 11, Тел.: (0 365) 221-29-14.

Автореферат диссертации разослан «8» октябрь 2020 года.

(Реестр протокола рассылки № 1 от 4 октября 2020 года).

  
**М.К.Абузалова**  
Председатель Ученого совета по присуждению ученых степеней,  
доктор филологических наук, доцент

**Г.И.Топрова**  
Ученый секретарь Ученого совета по присуждению ученых степеней,  
кандидат филологических наук,  
доцент

**Д.З.Ражабов**  
Председатель Ученого семинара при Ученом совете по присуждению ученых степеней, доктор филологических наук, доцент

## ВВЕДЕНИЕ (аннотация докторской диссертации)

**Актуальность и востребованность темы диссертации.** В мировой лингвистике процесс изучения проблем компьютерной и корпусной лингвистики начался в 40-е годы XX века, ускорился в 60-е годы, и в начале XXI века появились языковые корпуса из миллионов слов. Расширились возможности использования автоматического перевода, тезауруса, электронных словарей, были созданы их научно-теоретические базы и использованы на практике. Эти обновления проложили путь к появлению перспективных научных направлений, связанных с внедрением информационных технологий в лингвистику, что, в свою очередь, поставило на повестку дня необходимость изучения принципов разметки номинативных единиц в качестве материала корпуса.

В мировом языкознании к XXI веку активизировалось движение за научно-теоретическое изучение корпусной лингвистики. В этих условиях проблемы повышения качества автоматического перевода в направлении компьютерной лингвистики, теории лемминга и тегирования слов в каждом языке, алгоритма и создания лингвистического обеспечения, разработки программ анализа текста, тегирования, машинного перевода приобрели особую актуальность и значимость. В лингвистике, в частности в области компьютерной лингвистики, существует необходимость разработки семантического тегирования, моделей тегов, алгоритма тегирования, а также создания на этой основе автоматической программы семантического тегирования.

За годы независимости был проведен ряд работ в области компьютерной лингвистики по достижению автоматического перевода, искусственного интеллекта понимания и обработки узбекского языка. Поэтому среди всех научных направлений в лингвистике "...определение задачи поддержки научно-исследовательской работы по развитию государственного языка, осуществления международного сотрудничества в этой области"<sup>29</sup> указывает на важность проведения исследований, связанных с упрочнением позиций узбекского языка в системе Интернета. Внимание к государственному языку в нашей стране возросло до уровня одного из приоритетных направлений духовности. Поэтому наряду с сохранением, обогащением и повышением эффективности практического использования нашего родного языка, актуальным вопросом стало достижение широкого применения узбекского языка в современной информационно-коммуникационной системе. Учитывая тот факт, что корпусная лингвистика является перспективным научным направлением, создание Национального корпуса узбекского языка, изучение таких вопросов, как формирование лингвистических моделей на основе

---

<sup>29</sup> Указ Президента Республики Узбекистан УП-5850 от 21 октября 2019 года « О мерах по кардинальному повышению роли и авторитета узбекского языка в качестве государственного языка» . // [www.xabar.uz](http://www.xabar.uz)

современных научных принципов, являются актуальными задачами, стоящими перед нашей наукой.

Настоящее диссертационное исследование в определенной степени служит реализации задач, определенных в Указе Президента Республики Узбекистан УП-4997 от 13 мая 2016 года «Об организации Ташкентского государственного университета узбекского языка и литературы», Постановлениях ПП-4947 от 7 февраля 2017 г. «О стратегии действий по дальнейшему развитию Республики Узбекистан», ПП-2789 от 17 февраля 2017 года «О мерах по дальнейшему совершенствованию деятельности академии наук, организации, управления и финансирования научно-исследовательской деятельности», УП -5850 от 21 октября 2019 года «Указ Президента Республики Узбекистан «О мерах по кардинальному повышению роли и авторитета узбекского языка в качестве государственного языка», Постановлении Кабинета Министров Республики Узбекистан ПКМ -984 от 12 декабря 2019 года «Об утверждении Положения Департамента по развитию государственного языка», а также в других нормативно-правовых актах, касающихся данной сферы деятельности.

**Соответствие исследования приоритетным направлениям развития науки и технологий Республики Узбекистан.** Диссертационное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий Республики Узбекистан: 1. «Формирование системы инновационных идей и пути их реализации по социальному, правовому, экономическому, культурному, духовно-просветительскому развитию информационного общества и демократического государства».

**Степень изученности проблемы.** В мировой лингвистике целенаправленные исследования в области корпусов впервые были проведены в 40-е годы XX века Блумфилдом, Фризом и Бонджерсом<sup>30</sup>; Н.Фрэнсис и Г.Кучера впервые разработали принципы составления корпусов<sup>31</sup>. В русском языкознании В. П. Захаров, А.Б.Кутузов, Е.В.Недошивина, В.В.Риков, В.Плунгян проводили исследования о принципах построения корпуса, его типах, принципах составления и тегирования корпуса<sup>32</sup>. Проблемы лексико-семантического тегирования корпусных единиц были

---

<sup>30</sup> Блумфилд Л. Язык. – Москва: Прогресс, 1968. – 608 с.; Fries Ch.C. The structure of English. An introduction to the construction of English sentences. – New York: Harcourt, 1952. -304 p.; Bongers H. The history and principles of Vocabulary control. – Woerden: WOCOPI, 1947. - 442 p.

<sup>31</sup> Фрэнсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – Москва, 1967.; Синклер Д. Предисловие к книге “Как использовать корпуса в преподавании иностранного языка”/ Д. Синклер [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/corpora-info.html>, свободный.

<sup>32</sup> Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – 69 с. – С. 27.; Кутузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf; Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург, 2006.–26 с.; Риков В.В. Курс лекций по корпусной лингвистике. URL: <http://rykov-cl.narod.ru/c.html>; Плунгян В. Зачем мы делаем Национальный корпус русского языка? “Отечественные записки” 2005, №2. [http://magazines.russ.ru/oz/2005/2/2005\\_2\\_20-pr.html](http://magazines.russ.ru/oz/2005/2/2005_2_20-pr.html)

предметом исследований Г.И.Кустовой, О.Н.Ляшевской, Е.В.Падучевой, Е.В.Рахилиной<sup>33</sup>, Д.Апресьяна, Л.Л.Иомдина, Е.В.Санникова и В.Г.Сизова.<sup>34</sup>

В узбекском языкознании проведены соответствующие исследования по компьютерной лингвистике, лексической обработке текста и лингвистическому анализу. В качестве подобных исследований, считаем уместным, отметить научные труды А.Пулатова, С.Мухаммедова, М.Айимбетова, С.Мухамедовой, С.Каримова, Г.Джуманазаровой, А.Бабанарова, Д.Ербоевой, Н.Абдурахмоновой и А.Норова<sup>35</sup>. Данные исследования были актуальны тем, что с помощью текстового инновационного подхода – достижений компьютерной лингвистики удалось дать важные рекомендации по современным методам лексикографических и лингвостатистических исследований. В узбекском языкознании вопросы корпуса, его особенности, социальная, лексическая, образовательная значимость, типы корпуса, лингвистическая ценность авторского корпуса, лингвистические основы создания авторского корпуса узбекского языка изучены Ш.М.Хамроевой<sup>36</sup>, база слов-синонимов для корпусов узбекского языка исследована А.А.Эшмуминовым<sup>37</sup>. В процессе подготовки диссертации были тщательно изучены исследования вышеупомянутых лингвистов, прокомментированы и использованы в данном исследовании.

**Связь исследования с планами научно-исследовательских работ высшего учебного заведения, где выполнена диссертация.** Диссертация выполнена в соответствии с научно-исследовательскими планами Бухарского государственного университета в рамках научно-исследовательской работы на тему «Проблемы исследования языковых, личностных и общественных отношений в узбекском языкознании».

---

<sup>33</sup> Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Семантическая разметка лексики в национальном корпусе русского языка: принципы, проблемы, перспективы \\\ (Электрон ресурс): <http://ruscorpora.ru/sbornik2005/10kustova.pdf>;

<sup>34</sup>Апресьян Ю.Д., Иомдин Л.Л., Санников А.В., Сизов В.Г. Семантическая разметка в глубоко аннотированном корпусе русского языка // Труды международной конференции “Корпусная лингвистика – 2004”. – Санкт-Петербург: Издательство Санкт-Петербургского университета, 2004. – С.41-54.

<sup>35</sup> Мухаммедов С. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дисс... канд. фил. наук. – Ташкент, 1980.; Бабанаров А. Разработка принципов построения словарного обеспечения турецко-русского машинного перевода: Автореф. дисс... канд. фил. наук. – Ленинград, 1981.; Айымбетов М.К. Опыт лингвостатистического анализа лексики и морфологии каракалпакского публицистического текста: Автореф. дисс... канд. фил. наук. – Ташкент, 1987.; Каримов С., Каршиев А., Исроилова Г. Абдулла Қаҳҳор асарлари тилининг луғати. Алфавитли луғат. Частотали луғат. Терс луғат. – Тошкент, 2007.; Ризаев С. Ўзбек тилининг лингвостатистик тадқиқи: Фил. фан. док. дис... автореф. – Тошкент, 2008.; Мухаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. –Тошкент, 2006.; Ўринбоева Д.Б. Ўзбек фольклори матнларининг лингвостатистик тадқиқи. – Тошкент: Фан, 2010.; Жуманазарова Г.У. Фозил Йўлдош ўғли дostonлари тилининг лингвопоэтикаси: Фил. фан. док. дисс... автореф. – Тошкент, 2017.; Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (содда гаплар мисолида): Фил.фан.бўйича фалсафа доктори (PhD)...дисс. автореф. – Тошкент, 2018.; Пулатов А. Компьютер лингвистикаси. – Тошкент: Akadernashr, 2011.; Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.

<sup>36</sup> Хамроева Ш.М. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Фил.фан.бўйича фалсафа доктори (PhD)...дисс. – Бухоро, 2018. – 250 б.

<sup>37</sup>Эшмуминов А.А.Ўзбек тили миллий корпуси синоним сўзлар базасини яратиш асослари: Фил.фан.бўйича фалсафа доктори (PhD)...дисс. – Термиз, 2019. – 140 б.

**Целью исследования** является изучение значимости лексико-семантических связей между номинативными единицами узбекского языка в формировании корпуса; разработка моделей семантического тегирования номинативных единиц и лингвистического обеспечения лексико-семантического тегирования единиц узбекского языка в целях создания корпуса.

**Задачи исследования:**

исследовать значение теории лексико-семантического поля в формировании корпуса узбекского языка и место лексико-семантических отношений в тегировании номинативных единиц;

определить возможности использования в корпусной лингвистике достижений узбекской лексикографии;

изучить взгляды касательно теории лексико-семантического тегирования;

определить алгоритм семантического тегирования;

разработать модели тегов семантических полей узбекского языка;

выявить проблемы тегирования омонимов и слов множественного числа в узбекском языке и разработать оптимальные способы тегирования таких единиц;

семантически протегировать текст рассказа Абдуллы Каххара «Гранат» («Анор») и обобщить полученные результаты.

В качестве **объекта исследования** были выбраны номинативные единицы узбекского языка.

**Предметом исследования** являются лингвистические основы и модели лексико-семантического тегирования.

**Методы исследования.** При освещении темы диссертации использованы классификационный, описательный, сравнительный методы, а также метод статистического анализа.

**Научная новизна исследования** заключается в следующем:

исследована значимость теории лексико-семантического поля в узбекском языке в формировании корпуса, обоснована роль лексико-семантических отношений в семантическом тегировании номинативных единиц узбекского языка;

доказано, что взаимосвязь между узбекской лексикографией и корпусной лингвистикой служит лингвистическим обеспечением для семантического тегирования языкового корпуса лексикографических материалов;

определен алгоритм создания категорий семантических тегов, констант и моделей операторных тегов для номинативных единиц узбекского языка;

обоснованы теоретические основы разметки многозначных слов в узбекском языке и технология семантической фильтрации для разметки многозначных слов.

### **Практические результаты исследования:**

определены семантические модели тегов на основе семантических полей узбекского языка;

выявлены проблемы тегирования омонимов в узбекском языке и доказаны пути их решения;

разработан алгоритм составления категорий и моделей семантических тегов;

обоснованы методы семантического тегирования и лингвистического моделирования номинативных единиц узбекского языка;

семантически протегированы номинативные единицы в тексте рассказа Абдуллы Каххара «Гранат» и определены результаты на практике.

**Достоверность результатов исследования** объясняется тем, что изученные материалы помогают сделать выводы, основанные на характере узбекского языка, их обосновании, методологическом совершенствовании, опираются на проверенные практикой источники при создании принципов семантического тегирования номинативных единиц узбекского языка.

**Научная и практическая значимость полученных результатов исследования.** Исследование служит научно-теоретическим источником в разработке теоретических основ создания корпуса узбекского языка, в создании научных исследований по направлению компьютерной лингвистики.

**Практическая значимость исследования** заключается в его использовании в качестве источника при составлении учебных программ, планов и изложении тем в процессе преподавания таких дисциплин раздела прикладной филологии, как компьютерная лингвистика, корпусная лингвистика.

**Внедрение результатов исследования.** В исследовании на основе описания алгоритма создания моделей лексико-семантических тегов, лексико-семантического тегирования научно обоснованных номинативных единиц узбекского языка:

выводы касательно значимости теории лексико-семантического поля в формировании корпуса узбекского языка и места лексико-семантических отношений в тегировании номинативных единиц применены в реализации фундаментального научно-исследовательского проекта F1-ХТ-0-19919 на тему: «Узбекская мифология и ее роль в развитии художественного мышления» (справка № 89-03-2673 Министерства высшего и среднего специального образования Республики Узбекистан от 3 июля 2019 года). На основе научных результатов представлено отношение узбекской лексикографии и корпусной лингвистики, изучены и обобщены подходы к теории лексико-семантического тегирования;

материалы, раскрывающие теоретические основы разработки моделей семантических тегов, тегирования омонимов и многозначных слов в узбекском языке на основе семантических полей данного языка, использованы в реализации проекта практического гранта ИТД-1-07 на тему: «Создание учебника, учебно-методического комплекса и их электронного варианта по

учебной дисциплине «Основы языкознания»» (справка № 89-03-2152 Министерства высшего и среднего специального образования Республики Узбекистан от 19 июня 2020 года). Использование результатов исследования послужило разработке теоретических основ создания корпусов узбекского языка;

материалы, демонстрирующие алгоритм создания семантических категорий тегов и моделей, были использованы для создания сценариев телевизионных программ «Маърифат ёғдуси» и «Машварат» на телеканале "Махалля" Национальной телерадиокомпании Узбекистана (справка № 01-18-132 Национальной телерадиокомпании Узбекистана от 10 февраля 2020 года). В результате это способствовало возникновению в мировоззрении аудитории представлений о разработке семантических моделей тегов на основе семантических полей узбекского языка, заинтересованности молодых людей в приобретении знаний, а также совершенствованию их знаний узбекского языка через глобальную сеть Интернет;

**Апробация результатов исследования.** Результаты данного исследования обсуждены на 5 международных и 7 республиканских научно-практических конференциях.

**Публикация результатов исследования.** Основное содержание диссертации опубликовано в 20 работах. Из них 8 научных статей в научных изданиях, в том числе, 5 статей в республиканских и 3 статьи в зарубежных журналах, рекомендованных Высшей аттестационной комиссией Республики Узбекистан для публикации основных научных результатов докторских диссертаций.

**Структура и объём диссертации.** Диссертация состоит из введения, трёх глав, заключения, списка использованной литературы и приложения. Общий объём диссертации составляет 146 страниц.

## **ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ**

Во **введении** обоснованы актуальность и востребованность темы диссертации, определены степень изученности выбранной темы, цель и задачи, объект и предмет исследования. Сформулированы научная новизна, научная и практическая значимость работы, обоснована достоверность полученных результатов, приведены сведения о внедрении результатов исследования в практику, об апробации, результатах, опубликованных работах, охарактеризована структура диссертации.

В первой главе диссертации, озаглавленной «**Лексико-семантические отношения между номинативными единицами узбекского языка**», были изучены значимость теории лексико-семантического поля в формировании корпуса в узбекском языке, роль лексико-семантических отношений в тегировании номинативных единиц, а также взаимосвязь узбекской лексикографии и корпусной лингвистики. В первом параграфе главы, озаглавленном «**Значение теории лексико-семантического поля в формировании корпусов узбекского языка**», рассматриваются история



системного подхода к языку, использование результатов исследований по результатам синонимического ряда, лексико-семантической группы, лексико-семантического поля в формировании корпусных семантических тегов, значение ОСГ, ТГ и ЛСП в формировании корпусных семантических тегов. Язык – это система элементов, образующих единое целое. Значение каждого элемента системы вытекает из существования других элементов, одновременно составляющих систему<sup>38</sup>. В лингвистических концепциях, выдвинутых выдающимися учеными мировой лингвистики, такими как В. Гумбольдт, А. Шлейхер, Г. Штейнталь, В. Матезиус, В. Скаличка, Н.С. Трубецкой, Л. Ельмслев, Х. Ульдадь, Л. Блумфильд, Э. Сепир, Б. Уорф, З.С. Харрис, А. Мартинс, Л. Вейсгербер, Бодуэн де Куртэне, Ф.Фортунатов, можно встретить самые разнообразные проявления системно-структурного подхода к языку<sup>39</sup>. В русском языкознании лексическая система и структура в достаточной степени изучены<sup>40</sup>, заслуги узбекских ученых в подробном изучении лексики узбекского языка несравнимы. Своими фундаментальными исследованиями они не только заложили мощный фундамент для изучения лексики узбекского языка как системы, но и успешно смогли использовать данный метод на ранних его этапах, основали системную лексикологию узбекского языка<sup>41</sup>, а научные исследования группы ученых непосредственно посвящены изучению лексики с точки зрения системного языкознания<sup>42</sup>.

На основе теории лексико-семантического поля (ЛСП) межотраслевые исследовательские работы начались в основном в 70-80-е годы XX века. Этот этап характеризуется изучением лексики как целостной системы, состоящей из отношений определенных семантических элементов, входящих в определенную семантическую группу (ЛСГ).

<sup>38</sup> Соссюр Ф. Де. Труды по языкознанию. – Москва: Прогресс, 1977. – С.147.

<sup>39</sup> См: Основные направления структурализма. –Москва: Наука, 1964.; Звегинцев В.А. История языкознания XIX и XX веков в очерках и извлечениях. – ч. I. – Москва, 1964.; ч. II. – Москва, 1965.; Звегинцев В.А. Язык и лингвистическая теория. – Москва: Издательство МГУ, 1973.; Философские основы зарубежных направлений в языкознании. – Москва: Наука, 1977.; Новое в лингвистике. – Москва: Прогресс. Вып. I. 1960.; Вып. 2. 1962.; Вып. 3. 1963.

<sup>40</sup> См: Щерба Л.В. Языковая система и речевая деятельность. – Ленинград, 1974.; Виноградов В.В. Избранные труды. Лексикология и лексикография. – Москва: Наука, 1977.; Ахманова О.С. Очерки по общей и русской лексикологии. –Москва: Учпедгиз, 1957.; Уфимцева А.А. Опыт изучения лексики как системы (на материале английского языка). – Москва: Издательство АН, 1962.; Степанов Ю.С. Семиотика. –Москва:Радуга, 1983.; Шмелов Д.Н. Проблемы семантического анализа лексики. – Москва, 1973.; Новиков Л.А. Семантика русского языка. – Москва: Высшая школа, 1982.

<sup>41</sup> Ибрагимов С. Профессиональная лексика узбекского языка: Автореф. дис... д-ра фил. наук. – Ташкент, 1961.; Хожиев А. Узбек тили синонимларининг изоҳли луғати. –Тошкент: Ўқитувчи, 1974.; Раҳматуллаев Ш. Узбек тилининг изоҳли фразеологик луғати. –Тошкент: Ўқитувчи, 1978.; Кўчқортоев И. Сўз маъноси ва унинг валентлиги. –Тошкент: Фан, 1977.; Бегматов Э. Ҳозирги ўзбек адабий тилининг лексик қатламлари. –Тошкент: Фан, 1985.; Пинхасов Я. Ҳозирги ўзбек адабий тили (Лексикология ва фразеология). –Тошкент: Фан, 1969.; Дадабоев Х. Военная лексика в староузбекском языке: Автореф. дис.... канд. фил. наук. – Тошкент, 1981.; Бафоев Б. Кўҳна сўзлар тарихи. – Тошкент: Фан, 1991.

<sup>42</sup> Раҳматуллаев Ш. Семема – мустақил тил бирлиги // Ўзбек тили ва адабиёти. – 1984, 5-сон.; Кўчқортоев И. Сўз маъноси ва унинг валентлиги. –Тошкент: Фан, 1977.; Бегматов Э., Нейматов Х., Расулов Р. Лексик микросистема ва унинг тадқиқ методикаси (Систем лексикология тезислари) // Ўзбек тили ва адабиёти. – 1989, 6-сон. – Б.35- 40.; Нейматов Х., Расулов Р. Узбек тили систем лексикологияси асослари. – Тошкент: Ўқитувчи, 1995.; Расулов Р. Узбек тилида холат феъллари ва уларнинг облигатор валентликлари. – Тошкент: Фан, 1989.; Бозоров О. Узбек тилида даражаланиш: Фил. фанлари д-ри ... дис. автореф. – Тошкент, 1997.; Искандарова Ш. Узбек тили лексикасини мазмуний майдон сифатида ўрганиш (шахс микромайдони): Фил. фанлари д-ри ... дис. автореф. –Тошкент, 1999.; Сафарова Р.Г. Гипонимия в узбекском языке (на материале общеупотребительных зоонимов): Автореф. дис... канд. фил. наук. –Тошкент, 1990.; Орифжонов Ш. Узбек тилида луғавий градуонимия: Фил. фан. номзоди ... дис. автореф. – Тошкент, 1996.; Қиличев Б. Узбек тилида партонимия: Фил. фан. номзоди ... дис. автореф. – Тошкент, 1997.

Группировка слов в более масштабные круги – это тематическая группа (ТГ); эта группа представляет собой совокупность слов различных категорий, основанных на общности предмета. Этот набор формируется на основе различных: парадигматических и синтагматических отношений. Например, ТГ «спорт» (футбол, забивание гола, стадион, болельщик) или ТГ «торговля» (торговля, сделка, рынок, магазин, продавец, покупатель, продажа, продавать). ТГ включает в себя различные ОСГ. Например, ЛСГ торгового помещения (магазин, киоск, бутик, супермаркет, ларёк), синоним (владеть, покупать), антоним (дорогой – дешёвый), гипоним (магазин – гастроном) входят в состав ТГ «торговля».

По мнению лингвистов, эти единицы употребляются в речи в смешанном порядке. В семантическом тегировании языковых корпусов граница ТГ, ЛСТ и ЛСП ясна, в поисковых окнах эта иерархия должна быть расположена правильно. «Если ТГ представляет языковую картину мира и его фрагменты, то ЛСП обозначает грани и отношения смысла»<sup>43</sup>.

Два подхода к изучению лексики: семасиологический (от слова к понятию) и ономасиологический (от понятия к слову) подходы определяют основные направления работы со словом в языковом корпусе, изучение слова, языковые действия. Подход к слову на основе тематической группы, лексико-семантической группы, семантического поля дает возможность перейти от понятия к слову в корпусе. В мире в большинстве существующих корпусов с семантической аннотацией (разметкой, тегом) применяется принцип «от понятия к слову».<sup>44</sup> На наш взгляд, при семантическом тегировании корпусного материала очень важно придерживаться принципа от слова к понятию, от понятия к слову. В этом случае поиск единиц происходит семантически в иерархии ЛСГ > ТГ > ЛСП. Это поиск «от понятия к слову». В первом окне поиска отображаются ЛСП; при вводе в каждое ЛСП помещаются более мелкие ЛСП; каждый из них будет состоять из ТГ, а ТГ и ЛСГ. При семантическом тегировании корпусного материала и размещении ЛСМ в поисковом окне осуществляется в несколько этапов.

Во втором параграфе главы, озаглавленном «Роль лексико-семантических отношений в тегировании номинативных единиц», рассматриваются типы лексико-семантических отношений в узбекском языке, их изучение; анализируются проблемы создания комплекса семантических тегов на основе лексико-семантических отношений между единицами узбекского языка.

Лексические единицы могут существовать в языке, формируя различные парадигмы на основе различных семантических установок. Синонимы, антонимы, градуонимические, партонимические, гипонимические отношения

---

<sup>43</sup> См. ссылку: <https://studfiles.net/preview/5837469/page/4/>: (Рублева О. Л.. Лексикология современного русского языка. – Москва, 2016.)

<sup>44</sup> См. ссылку: (<http://ruscorpora.ru/corpora-sem.html>)

входят в данные типы языковых отношений<sup>45</sup>. Проведены широкомасштабные исследования причин объединения языковых единиц на основе законов диалектики, образования отдельных лексико-семантических групп, отличия от других членов внутри группы по определенным признакам. В источниках типы отношений между членами парадигмы в семантическом поле подразделяются на такие группы отношений, как синонимические, гипергипонимические, холо-меронимические, градуонимические, функционально-конфликтные<sup>46</sup>.

Известно, что языковой корпус является необходимым рабочим инструментом лингводидактической, лингвистической и других социальных сфер. В исследованиях, посвященных корпусу, уже не раз обосновано, что корпус является большой информационной базой в изучении слов<sup>47</sup>.

При семантической разметке принцип “от слова к понятию” представляет собой присоединение описаний значений слов, лексем, вступающих в лексико-семантические отношения: синонимов, антонимов, гипонимов/гиперонимов, градуонимов, холо/меронимов в качестве тега. Система семантических тегов включает в себя описания словоформ, лемм, их значений, приведенных в толковых словарях, а также описания существующих синонимов, градуонимов, гипонимов/гиперонимов, холо/меронимов, антонимов. Целесообразно включить в состав семантического тега такие признаки, как категориальный признак, *значение, синоним, антоним, слот, происхождение, омоним, стиль, пароним, вариант, период применения*. Можно сказать, при семантической разметке языковых единиц для корпуса заключения, теоретический материал, словари, полученные в результате исследования лексико-семантических отношений в лингвистике, служат лингвистической базой обеспечения.

В третьем параграфе главы под названием «Узбекская лексикография и корпусная лингвистика» раскрываются словарная статья и общие, различные аспекты семантических тегов, важность использования словарей узбекского языка в качестве корпусного материала, словари, которые могут служить источником информации для языкового корпуса, структура словарной статьи, а также рассмотрен их потенциал в разметке единицы корпуса. Создание

---

<sup>45</sup> Sayfullayeva R. va b. Hozirgi o`zbek adabiy tili. O`quv qo`llanma. – Toshkent : Fan va texnologiya, 2009. – стр.112.

<sup>46</sup> Собиров А. Ўзбек тилининг лексик сатҳини системалар системаси тамойили асосида тадқиқ этиш. – Тошкент: Маънавият, 2004. –стр.140.

<sup>47</sup> Для получения более полной информации об этом см.: Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – 48 с. – С. 27.; Кутузов А.Б. Корпусная лингвистика. – (Электронный ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электронный ресурс) – //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf; Национальный корпус русского языка // Вопросы языкознания. – Москва: Наука, 2006. – С. 149-155.; [http://wikipedia.org/wiki/ Корпусная лингвистика](http://wikipedia.org/wiki/Корпусная_лингвистика). Статья в интернет-энциклопедии Википедия; [www.ruscorpora.ru/](http://www.ruscorpora.ru/)Плунгян В. Зачем мы делаем корпусы? “Отечественные записки” 2005, №2. [http://magazines.russ.ru/oz/2005/2/2005\\_2\\_20-pr.html](http://magazines.russ.ru/oz/2005/2/2005_2_20-pr.html); Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Фил.фан.бўйича фалсафа доктори (PhD)...дис. – Бухоро, 2018. – 250 б.

электронных версий текстов словарей узбекского языка является первым этапом связывания словаря с корпусом. Систематичность в лексической системе не выделяется так ярко, как на других уровнях языка. Представлена возможность инвентаризации лексики с точки зрения качества и количества, а также проведения обширных исследований. Основным источником семантической разметки языковых единиц является Толковый словарь узбекского языка<sup>48</sup>. В словаре упоминается не всякое значение главного слова, а его современные общеупотребительные значения, доступные для большинства людей. Если общее слово является однозначным, то словарная статья интерпретирует это значение; если слово многозначное, то каждое значение оформляется отдельно, в определенном порядке с использованием арабских цифр темно-черного цвета (жирным шрифтом). При тегировании по принципу «от слова к понятию», обозначение значений слов осуществляется в порядке, указанном в словаре. Но при поиске слова «от понятия к слову» необходимо приложить значение множественного числа слов к соответствующим ЛСГ, ТГ и ЛСП, потому что если множественное число слов принадлежит одной ЛСГ с одним значением, то другой ЛСГ/ЛСП со вторым значением будет релевантен.

В словарной статье количество иллюстративных примеров ограничено; в корпусе, наоборот: можно сослаться на контекст тысяч примеров искомого слова. Связывание существующего словаря антонимов и синонимов с базой корпуса, при поиске слова в данных словарях на основе гиперссылки дают хороший результат. Поскольку словарь, отражающий гипер-гипонимические, холо-меронимические отношения, в узбекском языке еще не создан, лексического материала для семантического тегирования слова на основе этих параметров нет. При семантическом тегировании, опираясь на исследования в этой сфере, данное действие осуществляется самостоятельно, вручную.

Вторая глава исследования называется «**Теоретические основы и модели лексико-семантического тегирования номинативных единиц**». В первом параграфе главы «Взгляды на теорию лексико-семантической разметки» приведены общие сведения о семантической разметке, проанализировано изучение семантической разметки и его применения на практике.

Семантический тег – это комплекс специальных символов или комментариев, которые указывают на то, что слово или фраза в корпусе языка принадлежит к определенной семантической категории или даже к меньшей семантической группе (ЛСГ, семантическое поле и группа)<sup>49</sup>. Семантические теги охватывают определение значение/я слов/а, обстоятельства, определяют его принадлежность к тематической группе, ЛСГ, семантическому полю,

---

<sup>48</sup>См. Ўзбек тилининг изоҳли луғати (А.Мадвалиев таҳрири остида). 5 жилдли. –Тошкент: “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти, 2000-2006.

<sup>49</sup> Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с. – С. 48-49.

характеристику деривации, определение значения, классификацию слов, состав совокупности толкований, связанных со словом омонимов, синонимов. Наблюдение за принципами семантической разметки существующего национального корпуса показало, что к одному слову в корпусе прилагаются три различных признака (категория, лексико-семантическая характеристика и словообразовательное описание). Лексико-семантические теги (1) таксономия (принадлежность к одной лексеме) – соответствующий тег для существительных, прилагательных, глаголов, наречий; (2) мереология (признак, обозначающий «целый – часть», признак, представляющий «элемент – полосу/группу») – тег, представляющий предметные и не предметные номинативные единицы; (3) топология (топологический статус обозначаемого объекта) – тег, относящийся к чему-либо; (4) каузация, тег связанный с глаголами; (5) оценка – это предметные и не предметные единицы, сгруппированные по полю, связанному с прилагательным и наречием.

Не смотря на то, что семантическая разметка не имеет единой стандартной формы, в данном процессе используются коды, состоящие из сочетания букв и цифр или только цифр. Первая буква или цифра обозначает общее семантическое значение, следующий знак является индикатором малой семантической группы, которая содержит узкое, частное значение слова. Семантический тег – это не только слово, но и сочетание многих сочетаний в семантическую группу, при этом сочетания, выражающие одно значение в разных сочетаниях, кодируются одним символом. Семантическая разметка – это рабочий инструмент, позволяющий выполнять лингвистические действия в инструменте корпуса: наличие семантической аннотированной единицы в корпусе создает не только отдельное слово, но и возможность реагировать на поиск конструкции в случае соединения, запрошенного пользователем. Семантическая разметка является продолжением и расширением морфологической разметки; она сама по себе состоит из системы трех групп знаков: (1) единицы, образованной словообразовательными показателями; (2) знака, принадлежащего к лексико-семантическому признаку: собственное существительное, нарицательное существительное, притяжательное и относительное прилагательное; прямое и переносное значение; (3) частно-семантический характер: тематическая (таксономическая) группа, отражающая семантическое поле<sup>50</sup>. Семантическая разметка корпуса разделяется на два основных вида на основе сочетания компетенций программиста и лингвиста:

1) терминологическая разметка – этап именованного понятия в тексте;

---

<sup>50</sup> 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.

2) обеспечение тегом, обозначающим отношения между единицами. Здесь имеется в виду определение роли любой единицы в семантическом поле, её взаимосвязи с другой единицей.

А. А. Кретов разделяет теги на константы, операторы-классификаторы<sup>51</sup>. Например, *Sp t: constr. здание и сооружение (дом, чердак, мост)*. В системе данных тегов *Sp* – константа, а *t: constr* – классификатор.

В главе «Семантические категории тэгов в корпусной лингвистике: эксперимент и анализ» были исследованы различные подходы, взгляды на семантическую систему тегов; группы тегов в семантической системе тегов. В корпусной лингвистике был проведен ряд исследований по решению проблемы семантического увеличения, его системы тегов, категорий тегов, проблем семантического тегирования, многозначности и омонимии<sup>52</sup>. В.В.Куканова на вопрос «Зачем вам нужна семантическая разметка?» отвечает словами Гиндина: «К описанию языка можно подойти с двух сторон: от формы к содержанию, а также от содержания к форме. Первый подход показывает чёткое, полное описание всех существующих значений той или иной языковой единицы, контекста, в котором она применяется. Вторым подходом, напротив, направлен на то, чтобы свести все формы, которые могут представлять один и тот же смысл, к одной точке, в которой собран определенный набор значений. Оба подхода одинаково актуальны для лингвистики и языкового образования. Учитель грамматики и словаря изучает родной язык (иностраный язык), при этом чувствует необходимость в двух различных подходах. Если встречается незнакомое слово или конструкция, то обращаются к словарю/грамматике; если человек сам должен о чем-то написать или говорить, то ищет слово/обозначение: он знает, что сказать, чувствует, но не может найти нужное слово/грамматическую форму»<sup>53</sup>.

---

<sup>51</sup> Кретов А.А. Анализ семантических помет в НКРЯ // <http://ruscorpora.ru/sbornik2008/11.pdf>.

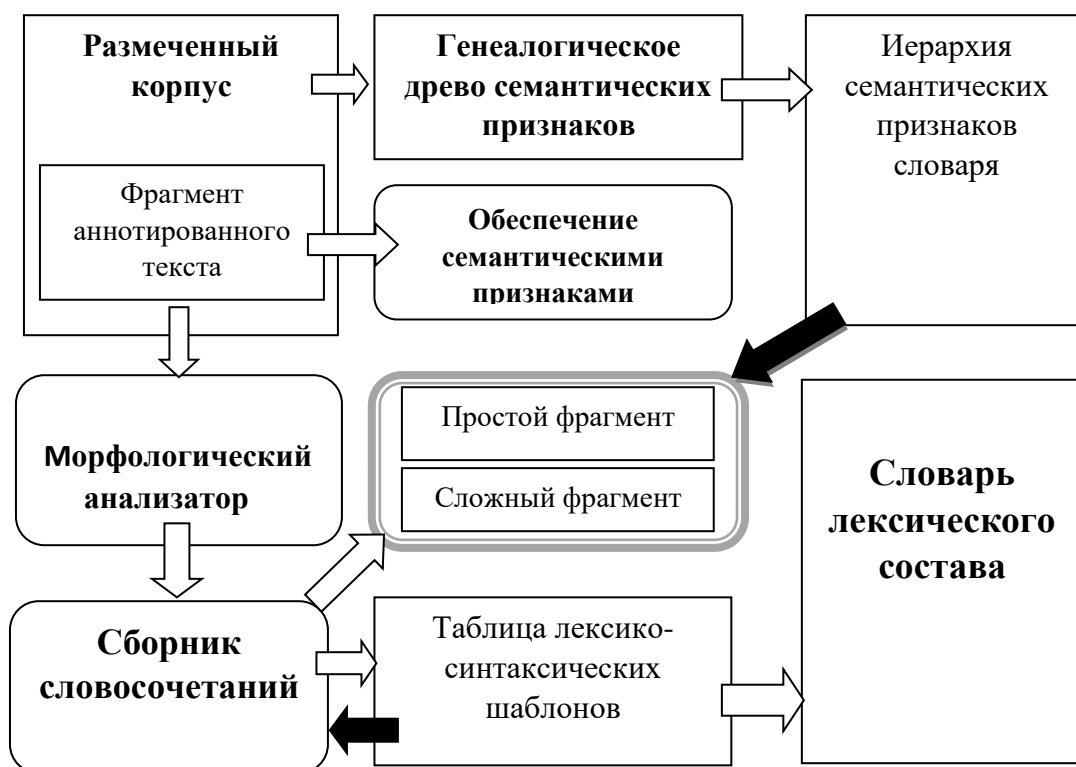
<sup>52</sup> Куканова В. В. Принципы семантической разметки национального корпуса калмыцкого языка // [http://kalmcorp.ru/sites/default/files/kukanova\\_25.pdf](http://kalmcorp.ru/sites/default/files/kukanova_25.pdf); Кретов А.А. Анализ семантических помет в НКРЯ // <http://ruscorpora.ru/sbornik2008/11.pdf>; 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.; 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.). Санкт-Петербург: Санкт-Петербургский гос.университет, Факультет филологии и искусств, 2008. – С.276-284.; 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». – Москва, 2006. –С.445-450.; 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. Санкт-Петербург: Нестор-История, 2009. – С.215-239.; 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.). Санкт-Петербург: Санкт-Петербургский гос.университет, Факультет филологии и искусств, 2008. – С. 276-284.

<sup>53</sup> Куканова В.В. Принципы семантической разметки национального корпуса калмыцкого языка // [http://kalmcorp.ru/sites/default/files/kukanova\\_25.pdf](http://kalmcorp.ru/sites/default/files/kukanova_25.pdf)

Действительно, семантическая разметка очень важна для изучения лексики языка: она открывает исследователю/пользователю/читателю широкий путь для наблюдения возможностей словосочетания, синтаксического построения. Например, есть некоторые прилагательные, которые могут образовывать семантическую валентность только с определенными существительными. Генерация таких запросов на интерфейсе показывает валентность слова через массу примеров.

Если лексическая единица имеет несколько значений в зависимости от контекста, она включается в требуемую группу в соответствии с её применением в контексте; эти теги также принадлежат к одному комплексу; отделены от других тегов точкой с запятой. Первый комментарий является основным, а комментарий после точки с запятой – значением в контексте. Увеличение осуществляется как в категории слов, так и в алфавитном порядке. Первый подход используется для того, чтобы осветить лексико-грамматический признак единицы, указать, к какой категории она относится. Если лексическая единица обозначает одно, независимо от контекста, то ставится знак постоянства, если значение меняется в зависимости от контекста, то знак изменчивости.

Система семантической разметки наряду с особенностями системы общелингвистической разметки, она также имеет несколько частных признаков. Аналогичная сторона других систем лингвистической разметки заключается в том, что она основана на лингвистическом обеспечении или базе. Специфическая сторона семантической разметки заключается в наличии двух типов аннотирования: фасетного и древесного метода. На практике один или два из этих двух методов, в зависимости от характеристик единицы, используются в смешанном виде. Если фасетный метод требует последовательного толкования единицы, то генеалогический метод требует обозначения группы, класса, области, к которой принадлежит единица. В первом случае она идет от слова к значению, во втором – от значения к слову. Это свидетельствует о валидности принципа слова —> значение, значение —> слово, как в процессе работы специалиста с разметкой, так и в процессе поиска, что отражается в интерфейсе. Иллюстрация составляющих средств разметки и их взаимоотношений приведена в следующей таблице:



В третьем разделе главы, озаглавленной «Алгоритм построения системы семантической разметки и моделей», обсуждаются проблемы системы семантической разметки, модели, многозначность и фильтр в системе семантической разметки.

Для различения многозначности необходим большой контекст: говорящий и слушающий не чувствуют такой множественности в речи. Корпусная разметка обычно срабатывает при определении многозначности в зависимости от присоединения единицы к соседнему слову. Для создания такого алгоритма лингвист должен ввести в базу лингвистического обеспечения набор правил (все типичные и редкие случаи). В системе разметки обозначение разряда единицы различается в зависимости от способности слова соединиться с названием предмета/личности: прилагательное *лёгкий* означает «не тяжелый», когда что-то образует сочетание с названием предмета. И следующий разряд этой единицы ограничивается таксономией непредметных названий. Один из его фильтров будет иметь следующий вид:

Слово	Контекст	Итоговое значение
<i>Лёгкий</i>	+ наименование предмета	Sem= разряд «признак- свойство» таксономический класс: физическое свойство: вес

Слово «голый» имеет два значения. Сравните: голый 1-голый человек; голый 2-голая мысль.

Каждое значение единства ограничено в своей связи с таксономическим классом существительного. Семантическая разметка может иметь следующий вид:

Слово	Контекст	Итоговое значение
-------	----------	-------------------



<i>Гольй</i>	+ личность	Sem= разряд «признак- свойство» таксономический класс: физическое состояние
<i>Гольй</i>	+ место	Sem= разряд «признак- свойство» таксономический класс: внешний вид
<i>Гольй</i>	+ наименование предмета	Sem= разряд «признак- свойство» таксономический класс: физическое свойство

Для фактического подтверждения нашего мнения мы проанализируем слово «холодный». Эта единица имеет такие значения, как холодный ветер (низкая температура), холодный цвет (цвет без красок), холодные глаза (психологическое состояние человека). Семантическая валентность может наблюдаться в следующем контексте:

<b>Слово</b>	<b>Контекст</b>	<b>Итоговое значение</b>
ХОЛОДНЫЙ	+ явление природы + время	Sem= разряд «признак- свойство» таксономический класс: физическое свойство: температура
ХОЛОДНЫЙ	+ цвет	Sem= разряд «признак- свойство» таксономический класс: физическое свойство: цвет-окраска
ХОЛОДНЫЙ	+ ментальная сфера + психическая сфера + качество присущее человеку + действия и поведение человека	Sem= разряд «признак- свойство» таксономический класс: качество присущее человеку

В качестве необходимого средства для корпусной семантической разметки необходимы словарь (1), семантический словарь, отражающий лексическую природу конкретного языка (2), лингвистическая модель – набор правил (3), система семантической разметки (4), дополнительный программный инструмент: фильтр способный различать множественность и омонимию (5). В процессе семантического тегирования необходима разработка специфических принципов разметки морфологической, лексической омонимии, универсальной лексики, входящей в состав сложного слова (составного термина), слова, отсутствующего в словаре, частичного фрагмента, буквенно-символической конструкции, так как данные единицы будут иметь особый характер в каждом языке.

Третья глава диссертации называется **«Проблемы создания лингвистического обеспечения лексико-семантических тегов корпусов узбекского языка»**. В первом разделе главы, озаглавленной «Значение семемы в семантическом тегировании языковых единиц», анализируется проблема семемы в семантическом тегировании языковых единиц, проблема подхода на основе словарной микросистемы в семантическом тегировании единиц узбекского языка. У Г.Глисона есть три компонента языка: структура выражения, структура содержания и словарь. Если единицей структуры выражения является морфема, то единицей структуры значения является семема. Эта классификация Г.Глисона подходит для корпусной лингвистики в семантическом аспекте леммы. Для морфологического тегирования глагола в

корпусе языка, исходя из структуры его выражения, в семантическом тегировании должна быть полностью затронута структура его значения. Если структура содержания не будет полностью раскрыта, то не будет возможности в полной мере продемонстрировать красоту нашего языка.

Установление границы поля также важно в семантическом соприкосновении корпусных единиц. Тот факт, что многозначные слова могут принадлежать только к одной области или они могут быть элементами нескольких областей, требует решения нескольких проблем при разметке глагола или леммы. Также на основе семантического фильтра можно будет определить, к какой области относится многозначное слово.

Во второй части главы, озаглавленной «Теоретические основы тегирования многозначных (полисемантических) слов в узбекском языке», исследуется подход к феномену многозначности в узбекском языке, лексическая интерпретация семантики и перенос их в корпус.

В корпусной лингвистике процесс семантического тегирования многозначных единиц должен разрабатываться с противоречивой точки зрения, с общих позиций. В результате корпусного поиска не все семы слова формируются на основе последовательности, так как результат поиска осуществляется на основе последовательности символов. Если языковой корпус не имеет хранилища данных семантической интерпретации, то результат формируется только на основе формы. Результат семантической разметки должен быть таким, чтобы семы можно было видеть в последовательности. Этого можно добиться только путем внесения словаря узбекского языка в лингвистическую базу языкового корпуса. Отсутствие такого словаря на узбекском языке указывает на необходимость создания новой модели словаря в узбекской лексикографии. Ведь создание узбекского языкового корпуса происходит пропорционально развитию узбекской лексикографии. В языковом корпусе отражается речевое воспроизведение языковых единиц: наблюдается, как слово употребляется в языке. Результатом поиска является пример, контекст, но семы не указаны. По нашему мнению, в семантическом оттенке многозначного слова необходимо различать его универсальное и терминологическое значение. В результате поиска необходимо сначала вывести в интерфейсе общее значение слова, а затем терминологическое значение. Здесь на основе соединительного словаря можно использовать формулы и модели, составленные обобщенным способом.

Какие результаты даёт лемма *человек* в языковом корпусе в ходе опроса? Эту лемму или буквально можно наблюдать с двух сторон. Если целью запроса является точно такой же «поиск слов», то можно выполнить базовый поиск или разговорный поиск. Языковой корпус находит и ссылается на сотни форм слова «человек», буквально столкнувшись с контекстами, доступными на базе. Пользователь не может различать значения лексемы «человек», может только определять контекст, в котором употребляется слово, способность к валентности, а также частоту его употребления. В толковом

словаре узбекского языка приведён пример 5 значений лексемы «человек», поясненный в отрывке, или требуется составить алгоритм, указывающий на проявление многозначности этого слова в нескольких контекстах.

**Человек** 1. Живое существо, человек, человек, обладающий способностью мыслить, говорить и работать, благодаря этому превосходит все остальные творения. *Первобытный человек.*

2. Определенный член человеческого общества, личность. Хороший человек. Честный человек. Настоящий мужчина. Он хороший человек.

3. Коллективное хозяйство, каждый из членов семьи; душа.

4. Слуга или сторонник богатого или чиновника.

5. Представитель другой общины или семьи.

Нами приведены сами комментарии, в языковом корпусе (в случае несовершенной разработки семантического тега) слово встречается без его полисемантического аспекта различения, и в данном случае извлечение сем слов из примеров представляет некоторые трудности. Пользователь, не имеющий опыта работы со словарём, не сможет выполнить данное действие. Является ли важным для пользователя какое значение слова употреблено в определенном тексте?

1. Пользователь может захотеть определить контекст, в котором слово применяется, с какими словами оно с большей вероятностью сочетается.

2. Пользователь может захотеть понаблюдать за семами слова, его многозначностью.

В первом случае последовательность контекстов, которые выходят в результате поиска в корпусе языка, сама решает проблему пользователя. Во втором случае такие слова должны быть специально обработаны специалистом по составлению корпуса. Эксперименты показывают, что комментарии, приведенные в ТСУЯ к многозначным словам, могут быть объединены через гиперссылку. Если слово имеет многозначную природу, то результат поиска должен быть выделен отдельным цветом. При выборе таких слов открывается ссылка в гиперссылке, пользователь может прочитать семы таких слов, их речевое применение, как если бы это был отрывок. До сих пор в корпусной лингвистике не было найдено другого решения для разграничения содержательных границ слова.

В третьем разделе главы «Проблемы тегирования омонимов в узбекском языке» рассматриваются вопросы тегирования единиц омонимов, разработки комплекса семантических тегов для единиц узбекского языка: константа и оператор/классификатор тегов.

Алгоритм фильтрации омолексем (снятие омонимии) является важным процессом. Такие формы, как *боғлар I – боғлар II*, *боғлар I – боғлар II*, *боғлар I – боғлар II* также сохраняют свою живучесть при получении любой грамматической формы, ее невозможно отличить от фильтра на основе грамматических форм: даже при поиске существительного места, даже при поиске единицы измерения или слова, обозначающего группу, ее можно увидеть в интерфейсе приложения.

Есть возможность с помощью фильтра различать омонимы, характерные для разных категорий. Например, прилагательное *дайди* (бродяга) относится к разряду прилагательных или глаголов. Для таких случаев фильтр можно сделать следующим образом. Поскольку оно встречается в речи в форме *дайдиб юрмоқ, дайдимайдиган*, то фильтр на основе относительной валентности можно определить следующим образом:

- 1) *дайди* + б/*майдиган* = глагол;
- 2) *дайди* + любое слово = прилагательное.

слово “*енг*” также может быть выделена на основе такого фильтра: *енг* I существительное. Та часть платья, которая прикрывает руку. *Енгли, енгсиз, енглик, енгча. С рукавами, без рукавов, рукавичка.*

*енг* II глагол. Выйти победителем, превзойти. *Сени иш енгмасин, сен ишни енг*<sup>54</sup>. *Не позволяй работе одолеть тебя, одолей работу сам*<sup>55</sup>.

- 3) *енг*+ли/сиз/лик/ча=существительное.;
- 4) *енг*+ма/гин/ди/май/гунча/гани/гач/ган= глагол.

Такие слова следует собирать из словарей и создавать семантическую фильтрующую базу. Даже фильтрация большого количества омонимов, которые относятся к двум категориям (например, существительное и прилагательное), может быть проблематичной. Если есть взаимная омонимия у частей речи существительного и глагола, то предложенный выше фильтр может дать результат, но если омонимия между существительным и прилагательным, то предложенный выше фильтр может оказаться менее эффективным. Например,

*ёпик* I прилагательное. В закрытом состоянии, запёртый. *Эшик, деразалар ёпик. Двери, окна закрыты.*

*ёпик* II существительное. От устига *ёпиладиган* мато, жул [попона]. *Сайислар ... отлар устидаги ёпикларни ва тўқимларни олиб, ҳалиги тол дарахти остига ёзганлар*<sup>56</sup>.

Когда слово «*ёпик*» использовано в качестве существительного, то принимает окончания принадлежности, падежа, множественного числа; слово, использованное как прилагательное, маркируется только суффиксом - *роқ* (сравнительная степень). Такое состояние можно отличить с помощью фильтра. Но если слово *закрытый* (*ёпик*) – существительное в именительном падеже, если наблюдается употребление прилагательного в положительной/превосходной степени, то фильтр не будет действовать.

- 1) *ёпик*+лар/ни/нинг/га(қа)/да/дан = существ.;
- 2) *ёпик*+роқ= прилаг.; аммо
- 3) *ёпик*+любое слово=?

<sup>54</sup>Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. – Тошкент: Ўқитувчи, 1984. – 215 б.

<sup>55</sup> Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. –Тошкент: Ўқитувчи, 1984. – стр. 215.

<sup>56</sup> Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. –Тошкент: Ўқитувчи, 1984. – стр. 61.

Итак, наиболее необходимым инструментом для выделения омонимов в тексте является морфосемантический фильтр. Специальная программа, выделяющая омоним – создание морфо-семантического фильтра, регулирует поиск омонимов в языковом корпусе. При создании такого фильтра существующий словарь омонимов служит информационным банком. Изучается несколько категорий омонимов, существующих в словаре, и для каждой формы требуется разработать индивидуально-общую схему фильтрации. Фильтрация категории омонимов должна быть решена в последующих исследованиях.

## **ЗАКЛЮЧЕНИЕ**

1. Лексический подход «от слова к понятию» и «от понятия к слову» в языковом корпусе определяет основные направления выполнения лингвистических действий со словом и систематизируется в корпусе на основе лемм ЛСГ, ЛСП. Однако, охватывая слово в синонимическом ряду, ЛСГ или ЛСП, он опирается на общую основу семантики, на систему значений каждой лексической единицы в группе, в которой весь лексический состав языка сливается в семантические поля общения, отражая совершенную, особую языковую картину мира для каждого языка. Кроме того, лингвистическое обеспечение требуется для семантической разметки номинативных единиц, так как набор семантических тегов будет состоять из тега, обозначающего принадлежность к форме слова, лемме, микроЛСГ, макроЛСГ, состоять из тега, относящегося к ТГ и ЛСП.

2. Толковый словарь отличается от тезауруса тем, что все значения слова разъясняются рядом со словами, входящими в лексико-семантическую связь, при этом в толковом словаре разъясняется только значение слова; хотя тезаурус направлен на раскрытие большего количества признаков слова, он отличается от корпуса простотой поиска, ограниченными возможностями. Корпусы, обеспечивают пользователя ссылкой на контекст применения слова, указывают на возможные варианты сочетания слова с другими единицами, но не раскрывают все его семантические признаки.

3. В корпусе языка (как и в словаре) не будет необходимости ссылаться на другое слово, потому что искусственный интеллект обладает большим объёмом памяти, в котором бумага не тратится впустую, всегда есть возможность расширения объёма, большой объём не вызывает никаких трудностей у пользователя, а миллионы слов, могут быть помещены в размеры всего нескольких мегабайтов. Если количество иллюстраций в словарной статье ограничено, то можно будет ссылаться на контекст тысяч примеров одного слова, запрашиваемого в корпусе.

4. Знак, обозначающий принадлежность слова/сочетания к определённой семантической категории, специфицирующий значение слова, а также совокупность интерпретаций являются семантической разметкой. Семантические теги корпуса охватывают специфику значения/значений слова, составление описаний, связанных с омонимичностью и синонимичностью

слова, категоризацию слова, определение его принадлежности к тематической группе, ЛСГ, семантическому полю, деривационные характеристики, признаки номинативного значения. Поскольку система семантической разметки состоит из разряда, лексико-семантической характеристики и деривационной характеристики, лексико-семантические теги группируются по таксономии, мереологии, топологии, каузации и оценочному полю.

5. Специфичность семантической разметки заключается в том, что применяется два метода аннотирования: фасетный и древесный. В зависимости от свойств единицы применяется один из методов или сочетание обоих. Фасетный метод предполагает описание, интерпретацию единицы в последовательном порядке, в то время как древесный метод требует обозначения группы, класса, поля, к которым принадлежит единица. 6. Для семантической разметки корпуса в качестве важного инструмента различают единицы, отражающие лексику определенного языка, такие, как: 1) словник, 2) семантический словарь, способный полностью интерпретировать лексику языка, 3) лингвистический модуль для реализации семантической разметки – комплекс правил, 4) система семантической разметки, 5) дополнительный программный инструмент: фильтр, позволяющий различать многозначность и омонимию. В процессе семантического тегирования необходима разработка специфических принципов разметки морфологической, лексической омонимии, универсальной лексики, входящей в состав сложного слова (составного термина), слова, отсутствующего в словаре, частичного фрагмента, буквенно-символической конструкции, так как данные единицы будут иметь особый характер в каждом языке.

7. Единицы корпуса маркируются операторным и константным тегами в зависимости от их принадлежности к семантическому полю, классу, группе. Потому что установление границ поля имеет важное значение при семантической разметке. Возможность многозначного слова выступить в качестве элемента одного или нескольких полей требует решения определенных задач при тегировании словоформы/леммы, так как принадлежность многозначного слова тому или иному полю определяется только на основе семантического фильтра.

8. Языковой корпус, отличается от бумажного словаря тем, что у него имеется возможность показать все значения лексической единицы, но бумажные словари являются основным сырьем для языкового корпуса. Посредством семантического тегирования необходимо добиться последовательного появления сем, а это может быть достигнуто путем введения присоединяемого словаря узбекского языка в лингвистическую базу языкового корпуса. Отсутствие такого словаря на узбекском языке указывает на необходимость создания новой модели словаря в узбекской лексикографии.

9. В мировой корпусной лингвистике выделяют методы тегирования омонимичных единиц, устранения омонимии в процессе автоматического чтения текста, основанные на 1) грамматической норме и 2) на статистике. Омонимия должна определяться отдельно по отношению к литературной речи

и разговорной речи, так как в словаре присутствуют в основном лексемы, относящиеся к современному узбекскому языку, слова, характерные для разговорной речи, включаются в ограниченном количестве. Но омонимы, используемые в литературной речи и разговорной речи в языковом корпусе, используются точно так же, как и омонимы, не отраженные в словаре, могут наблюдаться в языковом корпусе. Наиболее необходимым инструментом для выделения омонимов в тексте является морфо-семантический фильтр, создание морфо-семантического фильтра регулирует поиск омонимов в языковом корпусе.

10. Устойчивые теги можно рассматривать как название семантического поля: второстепенные теги являются операторами/классификаторами, обозначающими принадлежность к ЛСП, а также служат для уточнения значения слова; константные метки обозначают принадлежность к лексико-семантическому полю. Эксперименты показали, что выделенные ЛСП, ЛТГ могут являться константой или операторно-классификаторным тегом, принадлежащим к константе, в то время как в дальнейших исследованиях обогащение ряда данных тегов, совершенствование константы и системы операторно-классификаторных тегов представляется закономерным.

**SCIENTIFIC COUNCIL PhD.03/27.02.2020.Fil.72.09  
FOR AWARDING SCIENTIFIC DEGREES  
AT BUKHARA STATE UNIVERSITY**  

---

**BUKHARA STATE UNIVERSITY**

**AKHMEDOVA DILDORA BAHODIROVNA**

**LINGUISTIC BASES AND MODELS OF LEXICAL-SEMANTIC  
TAGGING OF NOMINATIVE UNITS FOR UZBEK LANGUAGE  
CORPORA**

**10.00.01 – Uzbek language**

**DISSERTATION ABSTRACT  
for the doctor of philosophy degree (PhD) in philological sciences**

**Bukhara – 2020**



The theme of the dissertation for doctor of philosophy (PhD) was registered at the Supreme Attestation Commission under the Cabinet of Ministers of the Republic of Uzbekistan under № B 2019.2.PhD/Fil777

The dissertation has been carried out at Bukhara State University.

The abstract of the dissertation is posted in three languages (Uzbek, Russian, English (resume)) on the Scientific Council website (buxdu.uz) and on the website of «ZiyoNet» Information and Educational portal www.ziynet.uz.

**Scientific supervisor:** Mengliev Bakhtiyor Rajabovich  
Doctor of Philological Sciences, Professor

**Official opponents:** Raupova Laylo Raximovna  
Doctor of Philological Sciences, Professor

Abjalova Manzura Abdurashetovna  
Doctor of Philosophy (PhD)

**Leading organization:** Karshi State University

The defense of dissertation will take place on «<sup>th</sup>20<sup>th</sup> of October 2020 at 10<sup>00</sup>» at the meeting of the Scientific Council for awarding scientific degrees **PhD.03/27.02.2020.Fil.72.09** at Bukhara State University. (Address: 200117, Bukhara city, Mukhammad Ikbol street, 11. Tel: (0365) 221-29-14; fax: (0365) 221-27-57; e-mail: buxdu\_rektor@buxdu.uz)

The dissertation could be reviewed in information-resource center of Bukhara State University (registration number 966). (Address: 200117, Bukhara city, Mukhammad Ikbol street, 11. Tel: (0365) 221-29-14).

Dissertation abstract was sent out on «<sup>th</sup>6<sup>th</sup> of October 20 20»  
(Mailing report number № 1 on «<sup>th</sup>7<sup>th</sup> of October 20 20»



**M.K. Abuzalova**  
Chairman of the Scientific Council for awarding scientific degrees, Doctor of Philological Sciences (DSc)

**G.I. Toirova**  
Scientific Secretary of the Scientific Council for awarding scientific degrees, Doctor of Philosophy (PhD)

**D.Z. Rajabov**  
Chairman of the Scientific Seminar at the Scientific Council for awarding scientific degrees, Doctor of Philological Sciences (DSc)

## **INTRODUCTION (abstract of PhD thesis)**

**The aim of the research work** is to study the importance of lexical-semantic relations between the units of the Uzbek language in the formation of corpus; development of models for semantic tagging of nominative units and elaboration of the corpus-based linguistic support for the lexical-semantic tagging of Uzbek language units.

**As the object of the research work** nominative units of the Uzbek language have been chosen.

### **Scientific novelty of the research work consists of the following:**

the importance of the Uzbek language lexical-semantic field theory in developing the corpus has been investigated and the role of lexical-semantic relations in semantic tagging of Uzbek nominative units have been grounded;

the function of the relationship of the Uzbek lexicography and corpus linguistics to serve as a linguistic support for the semantic tagging of lexicography materials linguistic corpus has been substantiated;

the algorithm for creating semantic tag categories as well as the constants and operational tags models for the Uzbek language nominative units has been defined;

the theoretical basis of tagging the Uzbek polysemantic words and the technology of semantic filtration for tagging polysemantic words have been proven.

**Implementation of the results of the dissertation.** Based on the description of the algorithm of the lexical-semantic tagging models of science-based Uzbek language corpus in the research:

conclusions on the role of lexical-semantic field theory in the Uzbek language to form the corpus and the role of lexical-semantic relations in the formation of nominative units have been used in the fundamental research project F1-XT-0-19919 "Uzbek mythology and its place in the development of artistic thinking" (reference No. 89-03-2673 of the Ministry of Higher and Secondary Special Education as of July 3, 2019). On the basis of scientific results, the relationship of the Uzbek lexicography and corpus linguistics has been shown, views on the theory of lexical-semantic tagging have been studied and generalized;

creating semantic tag models on the basis of semantic fields in the Uzbek language, theoretical bases of tagging of homonymous and ambiguous words in the Uzbek language have been used in the practical grant project ITD-1-07 (reference No. 89-03-2152 of the Ministry of Higher and Secondary Special Education of the Republic of Uzbekistan as of June 19, 2020). The application of the research results served to develop the theoretical basis for the creation of the Uzbek language corpus;

materials showing the algorithm for creating semantic tag categories and models have been used in the creation of scripts for the programs "Ma'rifat Yog'dusi" and "Mashvarat" of the National Television and Radio Company of Uzbekistan "Mahalla" (reference No. 01-18-132 of the National Television and Radio Company of Uzbekistan as of February 10, 2020). As a result, in the perceptions of audience it is observed that the development of semantic tag models based on the semantic fields of the Uzbek language, the interest of young people in learning and

improving their knowledge of the Uzbek language through the Internet have been established.

**The outline of the dissertation.** The dissertation consists of introduction, three chapters, conclusion, the list of used literature and appendix. The total volume of the work is 146 pages.

**ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ**  
**СПИСОК ОПУБЛИКОВАННЫХ РАБОТ**  
**LIST OF PUBLISHED WORKS**

**I бўлим (I часть; part I)**

1. Akhmedova D.B. Semantic labeling of language units// International Journal on Integrated Education. Indonesia. e-ISSN: 2620 3502 p-ISSN:2615 3785. Volume 3, Issue I, Jan 2020. – P.177-179. (SJIF 5.083)
2. Akhmedova D.B. Set of semantic tags for Uzbek language units: constants and operator/classifier// International Scientific Journal Theoretical & Applied Science. Philadelphia, USA. Impact Factor, 2409-0085(online) Issue: 02 Volume: 82 Published 29.02.2020. – P.177-179.
3. Ахмедова Д.Б. Вафо ва садоқат рамзи // Бухоро давлат университети илмий ахбороти. – Бухоро, 2015. №4.– Б. 110-111.(10.00.00; №1)
4. Ахмедова Д.Б. Лексик-семантик муносабатларнинг атов бирликларини теглашдаги ўрни// «Илм сарчашмалари». Урганч давлат университетининг илмий-методик журнали. – Урганч, 2019. №7.– Б. 107-112. (10.00.00; № 3)
5. Ахмедова Д.Б. Ўзбек тили бирликлари ўртасидаги лексик-семантик муносабат асосида семантик теглар мажмуини яратиш // Бухоро давлат университети илмий ахбороти. – Бухоро, 2019. №3.– Б. 115-121.(10.00.00; №1)
6. Ахмедова Д.Б. Корпус лингвистикасида семантик тег категорияси: тажриба ва таҳлил// Бухоро давлат университети илмий ахбороти. – Бухоро, 2020. №1. – Б.152-160.(10.00.00; №1)
7. Ахмедова Д.Б. Семантик разметка тизимида кўпмаънолилик ва филтёр// Бердақ номидаги Қорақалпоқ давлат университетининг ахборотномаси. – Нукус, 2020 . №1. – Б.202-205. (10.00.00; №12)
8. Ахмедова Д.Б. Семантик разметка тизимида тег гуруҳлари // «Oriental art and culture» scientific methodical journal. – Kokand, 2020. №3.– Б.440-444.
9. Akhmedova D.B. The approach to the phenomenon of polysemy in Uzbek// International consortium of education and culture research studies. International Conference on World Science and Education. Case of Asia. ISSN 2548 - 6160. Proceeding of the ICECRS. Vol 5 (2020) Conference of Management Quality University in the Industrial Revolution Era 4.0
10. Akhmedova D.B. The lexical microsystem based approach to semantic tagging of Uzbek language units// Pokonferensyjna Science Research Development. 25 01.2020, V2 Berlin. – P.257-260.
11. Ахмедова Д.Б. Семантик разметка тизими: умумий ва хусусий белгилар// «Ўзбек филологиясида матншунослик ва манбашунослик муаммолари» мавзусидаги III республика илмий-амалий анжумани. – Наманган, 12-13 июнь, 2020 йил. – Б.610-615.
12. Ахмедова Д.Б. Ўзбек тили бирликларини семантик теглашда луғавий микросистема асосида ёндашув// «Ўзбекистонда илмий-амалий тадқиқотлар» мавзусидаги республика миқёсидаги кўп тармоқли илмий масофавий онлайн конференция. №16. – Тошкент, 2 июнь, 2020 йил. – Б.284-286.

## II бўлим (II часть; part II)

13. Akhmedova D.B., Mengliyev B. Semantic Tag Categories in Corpus Linguistics: Experience and Examination// International Journal Of Recent Technology and Engineering (IJRTE) ISSN:2277-3878, Volume-8, Issue-3S, October-2019 (SCOPUS)

14. Ахмедова Д.Б., Қўшаева Н.Э. Муҳаммад Юсуф шеърини ва ёшлар маънавий тарбияси // «Баркамол авлодни миллий ва маънавий руҳда тарбиялаш муаммолари» республика илмий мақолалар тўплами. – Бухоро, 2014 йил. – Б. 304-306.

15. Ахмедова Д.Б. Ўзбек тилидаги луғатлар корпус материали сифатида// «Мактабгача ва бошланғич таълимнинг долзарб муаммолари» таълим муассасалари илмий-амалий конференция. – Бухоро, 18 апрель, 2019 йил. – Б. 186-189.

16. Ахмедова Д.Б. Ўзбек тили бирликлари ўртасидаги лексик-семантик муносабат асосида семантик теглар мажмуини яратиш// «Адабиётшунослик ва таржимашуносликнинг долзарб муаммолари: адабий жараён, қиёсий адабиётшунослик, услубшунослик ва тилшунослик масалалари» мавзусидаги республика илмий-амалий конференцияси материаллари тўплами.– Бухоро, 3-4 май, 2019 йил.– Б.261-263.

17. Ахмедова Д.Б., Ибрагимова З. Корпус материалларини теглашда ЛСГ, МГ, ва ЛСМнинг аҳамияти// «Замонавий таълимнинг барча босқичларида бола ҳуқуқлари соҳасидаги таълим фаолиятини такомиллаштириш» республика илмий-амалий конференцияси материаллар тўплами. – Бухоро, 28 ноябрь, 2019 йил. – Б. 31-34.

18. Ахмедова Д.Б., Ҳамидова И. Семантик разметка тизими ва моделларини тузиш алгоритми// «Тилшуносликдаги замонавий йўналишлар: муаммо ва ечимлар» мавзусидаги халқаро илмий-амалий 1 онлайн конференция.– Андижон, 28 январь, 2020 йил.– Б.692-695.

19. Ахмедова Д.Б. Семаларнинг лексикографик талқини ва уларнинг корпусларда берилиши// «Таълим тизимида чет тилларни ўрганишнинг замонавий муаммолари ва истиқболлари» мавзусидаги халқаро илмий-амалий анжуман. –Бухоро, 5-6 март, 2020 йил. – Б.732-735.

20. Ахмедова Д.Б. Ўзбек тили бирликлари учун семантик теглар мажмуи: константа ва оператор/классификатор// «Бошланғич синф ўқувчилари билимини баҳолашда халқаро тажрибалардан фойдаланиш: муаммо ва ечимлар» мавзусидаги республика илмий-амалий анжумани.–Бухоро, 14 март, 2020 йил. – Б.137-140.

21. Ахмедова Д.Б. Ўзбек тилидаги полисемантик сўзларни теглашнинг назарий асослари// «Замонавий филологияда интеграцион жараёнлар» мавзусидаги республика миқёсидаги онлайн илмий-амалий анжумани.– Бухоро, 25 апрель, 2020 йил.– Б.27-30.

22. Ахмедова Д.Б. Ўзбек тилидаги омонимларни теглаш муаммолари// «Замонавий таълимда рақамли тизимларни қўллаш: филология ва педагогика

соҳасида замонавий тенденциялар ва ривожланиш омиллари» мавзусидаги халқаро илмий-амалий масофавий конференция. – Тошкент, 1 май, 2020 йил.– Б. 344-346.