

ФАРҒОНА ДАВЛАТ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.03.30.12.2019.Fil.05.02 РАҚАМЛИ ИЛМИЙ КЕНГАШ
АЛИШЕР НАВОИЙ НОМИДАГИ ТОШКЕНТ ДАВЛАТ ЎЗБЕК ТИЛИ
ВА АДАБИЁТИ УНИВЕРСИТЕТИ

ГУЛЯМОВА ШАХНОЗА КАХРАМОНОВНА

ЎЗБЕК ТИЛИ СЕМАНТИК АНАЛИЗАТОРИНИНГ ЛИНГВИСТИК
АСОСЛАРИ

10.00.11 – Тил назарияси, амалий ва компьютер лингвистикаси

Филология фанлари доктори (DSc) диссертацияси
АВТОРЕФЕРАТИ

Докторлик (DSc) диссертацияси автореферати мундарижаси

Оглавление автореферата докторской (DSc) диссертации

Content of the abstract of doctoral (DSc) dissertation

Гулямова Шахноза Кахрамоновна

Ўзбек тили семантик анализаторининг лингвистик асослари.....3

Гулямова Шахноза Кахрамоновна

Лингвистические основы семантического анализатора узбекского языка.....37

Gulyamova Shakhnoza Kakhramonovna

Linguistic foundations of the semantic analyzer of the Uzbek language.....73

Эълон қилинган ишлар рўйхати

Список опубликованных работ

List of published works.....77

ФАРҒОНА ДАВЛАТ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.03.30.12.2019.Fil.05.02 РАҚАМЛИ ИЛМИЙ КЕНГАШ
АЛИШЕР НАВОИЙ НОМИДАГИ ТОШКЕНТ ДАВЛАТ ЎЗБЕК ТИЛИ
ВА АДАБИЁТИ УНИВЕРСИТЕТИ

ГУЛЯМОВА ШАХНОЗА КАХРАМОНОВНА

ЎЗБЕК ТИЛИ СЕМАНТИК АНАЛИЗАТОРИНИНГ ЛИНГВИСТИК
АСОСЛАРИ

10.00.11 – Тил назарияси, амалий ва компьютер лингвистикаси

Филология фанлари доктори (DSc) диссертацияси
АВТОРЕФЕРАТИ

Фан доктори (DSc) диссертацияси мавзуси **Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида В1021.1.DSc/FIL315** рақам билан рўйхатга олинган.

Докторлик диссертация Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетида бажарилган.

Диссертация автореферати уч тилда (Ўзбек, инглиз, рус (резюме)) Илмий кенгашнинг вебсаҳифасида (www.navoiy-uni.uz) ва «ZiyoNet» Ахборот тилим порталида (www.ziyounet.uz) жойлаштирилган.

Илмий маслаҳатчи:	Меиглиев Бахтиёр Ражабович филология фанлари доктори, профессор
Расмий оппонентлар:	Саментова Луиза Хамидовна филология фанлари доктори, профессор Улуҳов Носиржон Мухаммадалиевич филология фанлари доктори, профессор Шахабиддинова Шохида Ханимовна филология фанлари доктори, профессор

Етакчи ташкилот: Қарши давлат университети

Диссертация ҳимояси Фарғона давлат университети ҳузуридаги филология фанлари бўйича илмий даражалар берувчи DSc 03.30.12.2019 рақамли Илмий кенгашнинг 2022 йил «1» март соат 10⁰⁰ даги мажлисида бўлиб ўтади (Манзил: 100151, Фарғона шаҳри, Мураббийлар кўчаси, 19-уй. Тел: (373) 244-44-02; факс: (99871) 281-42-44.

Диссертация билан Фарғона давлат университетининг Ахборот-ресурс марказида танишиш мумкин (150 рақам билан рўйхатга олинган). Манзил: Манзил: 100151, Фарғона шаҳри, Мураббийлар кўчаси, 19-уй. Тел: (373) 244-44-02; факс: (9 9871) 281-42-44.

Диссертация автореферати 2022 йил 8 «феврал» кунин тарқатилди.
(2022 йил «9» «феврал» даги рақамли реестр баённомаси). «44»



Ш.М.Искаждарова
Илмий даражалар берувчи
илмий кенгаш раиси, ф.ф.д.,
профессор

М.Т.Зокиров
Илмий даражалар берувчи
илмий кенгаш илмий котиби,
ф.ф.и. доцент

А.Б.Мамажонов
Илмий даражалар берувчи
илмий кенгаш қошидаги илмий
семинар раиси, филол.ф.д.,
профессор

КИРИШ (докторлик (DSc) диссертацияси аннотацияси)

Диссертация мавзусининг долзарблиги ва зарурати. Жаҳон тилшунослигида 1930-1940 йилларда амалий тилшунослик, энг аввало, чет тилини ўқитиш жараёни, уни ўқитиш методикаси грамматиканинг ўзига хос хусусиятларини тасвирлаш орқали тушунилган. Россияда 1950 йиллардан бошлаб амалий тилшунослик компьютер технологиялари ва автоматик ахборотни қайта ишлаш тизимлари билан боғланган. 1960 йилларда махсус илмий йўналиш сифатида шаклланган амалий тилшунослик тилдан фойдаланишнинг амалий масалаларини ечишга қаратилган. Жумладан, жаҳон тилшунослигида XX асрнинг 60 йилларида пайдо бўлган корпус тилшунослиги нутқ амалиёти учун янги методологик асосни, тадқиқотнинг устувор йўналишларини яратди. Тил корпуси ва у билан боғлиқ бўлган фан, яъни корпус лингвистикаси XX асрнинг охири XXI аср бошида тилшунослик ва тил ўқитиш методикасига шиддат билан кириб келган янги йўналишдир. Тил корпусини яратиш ҳамда корпус лингвистикасини ривожлантириш тилшунослик, тил ўқитиш методикаси соҳаларидаги долзарб масалалардан бири ҳисобланади. Тил фаолиятининг замонавий шартлари нафақат нутқ намуналарини, балки алоҳида хронологик ҳолатлар, нутқ амалиётини тизимлаштиришни ўрганишни ўз ичига олади.

Дунё тилшунослигида XIX аср охири XX асрнинг бошларида корпуслар лингвистик тадқиқотлар ўтказиш ёки кўпинча амалий муаммоларни ечиш учун (масалан, тил бирликларининг частотасини ҳисоблаш учун) яратила бошланди. Корпус тилшунослиги учун компьютерларнинг ихтиро қилиниши ва кенг қўлланилиши билан ривожланишнинг янги босқичи бошланиб, яратилган корпуслар мавжудларидан нафақат сақлаш форматида, балки ҳажмларида ҳам фарқ қилишини таҳлил этиш долзарб масалага айланди. Компьютер лингвистикаси соҳасида турли мақсадли корпуслар яратиш, мавжуд корпусларни мукамаллаштириш, бойитиш, кенгайтириш кабилар ечимини кутаётган масалалардан биридир. Компьютер лингвистикасини компьютер воситаларидан фойдаланиш соҳаси сифатида дастурлар, маълумотларни қайта ишлаш технологияларини ташкил этиш ва муайян шароитларда тил фаолиятини моделлаштириш учун, шунингдек, тилшуносликда тилнинг компьютер моделлари ва унга оид фанлар доирасида қўллаш мумкин. Тил жуда мураккаб таълим бўлганлиги туфайли компьютер тилшунослигида автоматик матн таҳлили; автоматик матн синтези; автоматик луғатларни яратиш ва қўллаб-қувватлаш; автоматлаштирилган ахборот-қидирув тизимларини яратиш; машина таржимаси; автоматик тил ўрганиш тизимларини яратиш; лингвистик маълумотлар базаларини яратиш; назарий ва амалий тилшунослик масалаларини ечишнинг дастурий воситаларини ишлаб чиқиш каби турли йўналишлар ривожланиб бормоқда.

Ўзбекистонда мустақиллик йилларида компьютер ва корпус лингвистикаси бўйича санокли тадқиқотлар амалга оширилган. Бинобарин, “Бугунги глобаллашув даврида ҳар бир халқ, ҳар қайси мустақил давлат ўз

миллий манфаатларини таъминлаш, бу борада, аввало, ўз маданиятини, азалий кадриятларини, она тилини асраб-авайлаш ва ривожлантириш масаласига устувор аҳамият қаратиши табиий. Ўзбек тилининг халқимиз ижтимоий ҳаётида ва халқаро миқёсдаги обрў-эътиборини тубдан ошириш, униб-ўсиб келаётган ёшларимизни ватанпарварлик, миллий анъана ва кадриятларга садоқат, улуғ аждодларимизнинг бой меросига ворислик руҳида тарбиялаш, мамлакатимизда давлат тилини тўлақонли жорий этишни таъминлаш”¹ мақсадида корпус лингвистикаси истиқболли илмий йўналиш сифатида унинг назарий муаммоларини ечишни тақозо этади. Бу эса фанимиз олдида турган долзарб вазифалардан бири – ўзбек тили бирликлари автоматик таҳлилида омонимлик, кўп маънолилиқ ва полифункционалликни фарқлашнинг лингвистик асосларини яратишни кун тартибига қўяди.

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 февралдаги ПФ-4947-сон “Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида”ги Фармони, Ўзбекистон Республикаси Президентининг 2017 йил 17 февралдаги ПФ-2789-сон “Фанлар академияси фаолияти, илмий-тадқиқот ишларини ташкил этиш, бошқариш ва молиялаштиришни янада такомиллаштириш чора-тадбирлари тўғрисида”ги Қарори, 2019 йил 4 октябрдаги ПҚ-4479-сон “Ўзбекистон Республикасининг “Давлат тили ҳақида”ги Қонуни қабул қилинганлигининг ўттиз йиллигини кенг нишонлаш тўғрисида”ги қарори, Ўзбекистон Республикаси Президентининг 2019 йил 21 октябрдаги ПФ-5850-сон “Ўзбек тилининг давлат тили сифатидаги нуфузи ва мавқеини тубдан ошириш чора-тадбирлари тўғрисида”ги Фармони, 2020 йил 24 январдаги Ўзбекистон Республикаси Президентининг Олий Мажлисга Мурожаатномаси, Ўзбекистон Республикаси Президентининг 2020 йил 20 октябрдаги ПФ-6084-сон “Мамлакатимизда ўзбек тилини янада риволантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида”ги Фармони ҳамда бошқа меъриий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишда ушбу тадқиқот муайян даражада хизмат қилади.

Диссертациянинг Ўзбекистон Республикаси фан ва технологиялар тараққиёти устувор йўналишларига мослиги. Тадқиқот республика фан ва технологиялари ривожланишининг I. “Ахборотлашган жамият ва демократик давлатни ижтимоий, ҳуқуқий, иқтисодий, маданий, маънавий-маърифий ривожлантириш, инновацион иқтисодиётни ривожлантириш” устувор йўналишига мувофиқ бажарилган.

¹ Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг «Ўзбек тилининг давлат тили сифатидаги нуфузи ва мавқеини тубдан ошириш чора-тадбирлари тўғрисида»ги фармони // www.xabar.uz

Диссертация мавзуси бўйича хорижий илмий тадқиқотлар шарҳи².

Жаҳон тилшунослигида автоматлаштирилган ахборот-қидирув тизимлари лингвистик таъминотини ўрганишга йўналтирилган илмий изланишлар етакчи илмий марказлар ва олий таълим муассасаларда, жумладан, Oxford University (Буюк Британия), Cambridge University (Буюк Британия), Boston University, University of New York, University of Pennsylvania, University of California (АҚШ); Indiana University (АҚШ), The University of Sorbonne (Франция), Упсалла университети (Швеция), Kiofo University (Япония), Australian National University (Австралия), Истамбул техника университети (Туркия), Москва давлат университети (Россия), Москва давлат лингвистика университети (Россия), Санкт-Петербург давлат университети (Россия), Урал давлат педагогика университети, Харьков давлат педагогика университети (Украина), Baki Dövlət universiteti (Озарбайжон), Татаристон Республика Фанлар академияси амалий тилшунослик ва семиотика институти (Татаристон Республикаси, Россия), Қозон федерал университети (Татаристон Республикаси, Россия), В.Вернадский номидаги Таврия академияси Қрим федерал университети (Қрим Республикаси, Россия), Қирғизистон давлат техника университети (Қирғизистон давлат техника университети) шунингдек, Мирзо Улуғбек номидаги Ўзбекистон Миллий университети, Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университети, Ал-Хоразмий номидаги Тошкент ахборот технологиялари университети ва унинг филиаллари, Бухоро давлат университети, Термиз давлат университети (Ўзбекистон)да олиб борилмоқда.

Дунё тилшунослигида ахборот-қидирув тизимида семантик анализаторнинг ўрни, матнга автоматик ишлов бериш, матннинг автоматик анализ/синтез дастурлари, семантик анализаторлар тадқиқига оид олиб борилган изланишлардан қуйидаги илмий натижалар олинган: лингвистик анализатор тушунчаси, унинг турлари тавсифланган, семантик анализ ва синтезнинг матннинг автоматик таҳлилидаги аҳамияти ёритилган (Boston University, New York University, АҚШ; Уральский федеральный Университет, Россия); лингвистик парсер, синтактик парсернинг аҳамияти ва улардан фойдаланиш усуллари аниқланган (Уральский федеральный университет, Россия); матнга автоматик ишлов бериш технологиялари ўрганилган (University of Tokyo, Япония, Australian National University, Австралия, Oxford University Language Centre, Англия); Морфо-Анализатор, СегмМорф, Линтез, Пси-Ан каби таҳлил тизимларининг асослари ишлаб чиқилган (Харьков давлат педагогика университети, Украина; Қозон федерал университети (Россия, Татаристон Республикаси); матнга автоматик морфологик ва синтактик ишлов бериш модели ва дастурлари (Харьков давлат педагогика университети, Украина; Baki Dövlət universiteti,

² Диссертация мавзуси бўйича хорижий илмий тадқиқотлар шарҳи google.scholar.com., cheloveknauka.com, dslib.net, reseachgate. www.kaznu.kz, bloomsbury.com, www.utexas.edu, english.pku.edu.cn, www.lang.ox.ac.uk, www.bu.edu, americanenglish.state.gov, micros.uz, www.uv.es, www.zora.uzh.ch, cleo@snu.ac.kr, www.ras.ru, adwww.ecu.edu.au, www.sophia.ac.jp/eng; www.msu.ru; www.vsu.bu; www.education.ua/universities, www.navoiy-uni.uz.

Озарбайжон); туркий тиллар компьютер лингвистикасида табиий тилнинг компьютер ишлови, синтактик-семантик анализатор тузиш масалалари тадқиқ қилинган (Татаристон Республика Фанлар академияси амалий тилшунослик ва семиотика институти, Қозон федерал университети (Россия, Татаристон Республикаси), В.Вернадский номидаги Таврия академияси Қрим федерал университети (Россия, Қрим Республикаси), ўзбек-инглиз тили машина таржимасининг лингвистик таъминоти, ўзбек тили муаллифлик корпусини тузиш тамойиллари ва унинг лингвистик таъминоти, ўзбек тили бирликларини графематик таҳлил қилиш муаммолари, тил корпусининг лингвистик базасини тузиш тамойиллари ишлаб чиқилган; ўзбек тили атов бирликларини семантик теглашнинг назарий асослари, ўзбек-инглиз параллел корпуси инструментарийларини тузиш асослари, ўзбек тили таълимий корпусини тузишнинг лингвистик асослари, ўзбек миллий корпусида идиомалар базасини яратиш, ўзбек тили морфоанализатори ва унинг лингвистик таъминотини яратиш масаласи ишлаб чиқилган (Мирзо Улуғбек номидаги ЎзМУ, Алишер Навоий номидаги ТошДЎТАУ, Ал-Хоразмий номидаги ТАТУ, ТерДУ, БухДУ).

Жаҳон тилшунослигида табиий тилга автоматик ишлов бериш, анализаторлар яратиш бўйича, жумладан, ахборот-қидирув тизими таҳлил дастурлари, замонавий ахборот технологиялари воситаларида ишлов беришни йўлга қўйиш учун тил корпуслари, семантик анализатор ҳамда уларнинг лингвистик асосларини яратиш каби йўналишларда тадқиқотлар олиб борилмоқда.

Муаммонинг ўрганилганлик даражаси. Дунё компьютер тилшунослигида семантик анализ ва анализатор борасида бир қатор олимлар илмий изланиш олиб боришган. А.В.Тузов, М.В.Мозговой, А.В.Сокирко, Н.А.Счлаефер, А.В.Мочалова, Д.Ш.Сулейманов, А.Р.Гатиатуллин, Б.Ергеш, А.Шарипбай, Г.Бекманова, С.Липницки, Е.А.Каневский, К.К.Боярский³ каби

³ Тузов, В.А. Компьютерная семантика русского языка / В.А. Тузов. – СПб.: Изд-во С.-Петерб. ун-та, 2004. – 400 с.; Мозговой, М.В. Простая вопросно-ответная система на основе семантического анализатора русского языка / М.В. Мозговой // Вестник СПб университета. – 2005. – сер. 10. – вып. 1. – С. 116–122; Сокирко, А.В. Семантические словари в автоматической обработке текста (по материалам системы ДИАЛИНГ): дис. ... канд. техн. наук: 05.13.17 / Сокирко Алексей Викторович. – М., 2001. – 120 с.; Schlaefer, N. A Semantic Approach to Question Answering / N. Schlaefer . – Paperback, 2007. – 96 p.; Мочалова, А.В. Алгоритм семантического анализа текста, основанный на базовых семантических шаблонах с удалением / А.В. Мочалова // Научно-технический вестник информационных технологий, механики и оптики. – 2014.– № 5. – С. 126 –132.; Mathematical model of an ontological-semantic analyzer using basic ontological-semantic patterns / A.V.Mochalova, V.A.Mochalov // Lecture Notes in Artificial Intelligence, Proceedings of 15th Mexican International Conference on Artificial Intelligence. – 2016. – P. 53–66.; Семантический анализ. Некоторые проблемы автоматической обработки текста / А.В.Мочалова, В.П.Захаров, В.А.Кузнецов, В.А.Мочалов. – Петрозаводск: ПИН, 2016. – 60 с.; Семантический анализатор русскоязычного текста для вопросно-ответной системы // ДИССЕРТАЦИЯ на соискание ученой степени кандидата технических наук. – Петрозаводск, 2017. – 128 с.; Сулейманов Д.Ш., Гатиатуллин А.Р. Структурно-функциональная компьютерная модель татарских морфем. — Казань, 2003. — 220 с. – С. 152.; Yergesh B, Sharipbay A, Bekmanova G, Lipnitskii S. Sentiment analysis of kazakh phrases based on morphological rules // Известия Кыргызский государственный технический университет, теоретический и прикладной научно-технический журнал 2016 № 2(38). – С. 39-43.; Каневский Е.А., Боярский К.К. Семантико-синтаксический анализатор SemSin // Компьютерная лингвистика и интеллектуальные технологии. 2012; Боярский К.К., Каневский Е.А. Предсинтаксический модуль в анализаторе SemSin // Интернет и современное общество. Санкт-Петербург, 2013. С. 280–286.

олимпларнинг тадқиқотларини ана шундай ишлар сифатида қайд этиш мумкин.

Жаҳон ва туркий тилшуносликда морфологик/синтактик анализатор тузиш масалалари кенг тадқиқ этилган. Жумладан, А.В.Дыбо, А.К.Асирян, А.В.Шеймович, А.Р.Гатиатуллин, А.М.Баширов, П.В.Желтов, Н.А.Исраилова, П.С.Бакасова, М.Г.Мальковский, А.С.Старостин, Д.Н.Ёунгер, Е.А.Каневский, Н.В.Колпакова, Н.А.Исраилова, П.С.Бакасова, С.Бирд, Э.Клеин, Э.Лопер, М.Коробов, А.Д.Москвина, Д.Орлова, П.В.Паничева, О.А.Митрофанова, О.Н.Ляшевская, И.Астафьева, А.Бонч-Осмоловская, А.Гарейшина, Ю.Гришина, В.Дьячков, М.Ионов, А.Королева, М.Кудринский, А.Литягина, Е.Лучина, Е.Сидорова, С.Толдова, С.Савчук, С.Коваль ва К.К.Боярский каби бир қатор олимплар тадқиқотида батафсил ёритилган⁴.

Корпус лингвистикасида омонимлик муаммосини ечиш, омоним бирликларни теглаш ва матнни автоматик ўқиш жараёнида омонимияни бартараф этиш масаласи (“снятие омонимии”)га оид қатор тадқиқотлар вужудга келган. Жумладан, Г.И.Кустова, О.Н.Ляшевская, Е.В.Падучева,

⁴ Асирян А.К. Сравнение инструментов морфологической разметки / Intellectual potential of the XXI century 2017 / <http://www.sworld.education>; Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов тюркских языков // Филология и культура. Philology and culture. 2014. №2(36). – С. 20-26.; Гатиатуллин А.Р., Баширов А.М. Морфологический анализатор тюркских словоформ на базе структурнофункциональной модели тюркской морфемы // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». –Казань, 2017. – 327 с.; Гатиатуллин А. Р., Баширов А.М., Осипов Г.С., Смирнов И.В., Шелманов А.О. Методы лингвистического анализа текстов на татарском языке и их применение в поисковой системе Eхactus // Труды ИСА РАН. Том 66. 1/2016. – 18-25 с.; Желтов П.В. Разработка морфологического анализатора чувашского языка // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Казань, 2017. – 327 с.; Желтов П.В. Создание национального корпуса чувашского языка: проблемы и перспективы // Современные проблемы науки и образования. – 2015. – № 1-1.; URL: <http://www.science-education.ru/ru/article/view?id=19046> (дата обращения: 18.01.2021); Исраилова Н.А., Бакасова П.С. Морфологический анализатор кыргызского языка // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Казань, 2017. – 327 с.; Мальковский М.Г., Старостин А.С. Модель синтаксиса в системе морфо-синтаксического анализа «TREEON» // Тр. междунар. Конф. Диалог 2006. – М., 2006. – С. 481-492.; Younger D.H. Recognition and parsing of context-free languages in time n3. Information and Control 10 (2). – P.189-208.; Каневский Е.А., Колпакова Н.В. К вопросу построения морфологического анализатора // Компьютерная лингвистика и интеллектуальные технологии. 1999. Т. 2. С. 98–106; Исраилова Н.А., Бакасова П.С. Морфологический анализатор кыргызского языка // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. В 2-х томах. Т 2. – Казань: Издательство Академии наук Республики Татарстан, 2017. – 327 с.; Bird S., Klein E., Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. Beijing, 2009.; Korobov M. Morphological Analyzer and Generator for Russian and Ukrainian Languages // Analysis of Images, Social Networks and Texts: 4th International Conference, AIST 2015, Yekaterinburg, Russia, April 9–11, 2015. Communications in Computer and Information Science, Springer, 2015. P. 320–332.; Москвина А.Д., Орлова Д., Паничева П.В., Митрофанова О.А. Разработка ядра синтаксического анализатора для русского языка на основе библиотек NLTK // Компьютерная лингвистика и вычислительные онтологии. Труды XIX Международной объединенной научной конференции «Интернет и современное общество», Санкт-Петербург, 22–24 июня 2016 г. СПб, 2016. С. 44–54.; Ляшевская О.Н., Астафьева И., Бонч-Осмоловская А., Гарейшина А., Гришина Ю., Дьячков В., Ионов М., Королева А., Кудринский М., Литягина А., Лучина Е., Сидорова Е., Толдова С., Савчук С., Коваль С. Оценка методов автоматического анализа текста: морфологические парсеры русского языка // Компьютерная лингвистика и интеллектуальные технологии. 2010. № 9 (16). С. 318–326; Каневский Е.А., Боярский К.К. Морфолого-лексический анализатор и классификация текста // Прикладная лингвистика в науке и образовании. Материалы V международной научно-практической конференции. Санкт-Петербург, 2010. С. 157–163.

Е.В.Рахилина, Б.П.Кобрицов, Т.И.Резникова⁵, Б.П.Кобрицов⁶, В.В.Куканова⁷, А.А.Кретов⁸лар шу масалалар ечимига бағишланган қатор ишларни эълон қилишган.

Ўзбек тилшунослигида компьютер лингвистикаси, матнга лексикографик ишлов бериш борасида М.Маҳмудов, М.Айимбетов, С.Каримов, Г.Жуманазарова, А.Бабанаров, Д.Ўринбоева, А.Норов, С.Муҳаммедов, Н.Абдурахмоноваларнинг тадқиқотлари мавжуд⁹.

Ўзбек тили семантик анализаторини тузиш махсус тадқиқот предмети бўлмаган. Аммо ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш¹⁰, ўзбек-инглиз тили машина таржимасининг лингвистик таъминоти¹¹, ўзбек тили муаллифлик корпусини тузиш тамойиллари¹², ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи

⁵Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. М.: Индрик, 2005. – С. 155-174.; Ляшевская О.Н. Топологические классы имен в семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). СПб.: С.-Петербургский гос. университет, Факультет филологии и искусств, 2008. – С. 276-284.; Рахилина Е.В., Кобрицов Б.П., Кустова Г.И., Ляшевская О.Н., Шеманаева О.Ю. Многозначность как прикладная проблема: семантическая разметка в Национальном корпусе русского языка // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог-2006». – М., 2006. – С. 445-450; Рахилина Е.В., Кустова Г.И., Ляшевская О.Н., Резникова Т.И., Шеманова О.Ю. Задачи и принципы семантической разметки лексики в НКРЯ // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы / отв. ред. В.А.Плунгян. СПб.: Не-стор-История, 2009. – С.215-239.; Ляшевская О.Н. Топологические классы имен в семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). СПб.: С.-Петербургский гос. университет, Факультет филологии и искусств, 2008. – С. 276-284.

⁶ Кобрицов Б.П. Модели многозначности русской предметной лексики: глобальные и локальные правила разрешения омонимии. Автореф. канд. филол. наук.– М.: РГГУ, 2004.; Кобрицов Б.П., Ляшевская О.Н., Шеманаева О.Ю. Поверхностные фильтры для разрешения семантической омонимии в текстовом корпусе // Кобозева И.М., Нариньяни А.С., Селегей В.П. (ред.), Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции Диалог'2005. – М., 2005.

⁷ Куканова В.В. Принципы семантической разметки национального корпуса калмыцкого языка // http://kalmcorpورا.ru/sites/default/files/kukanova_25.pdf

⁸ Кретов А.А. Анализ семантических помет в НКРЯ // <http://ruscorpора.ru/sbornik2008/11.pdf>

⁹ Маҳмудов М.А., Пиотровская А.А., Садыков Т. Система машинного анализа и синтеза тюркской словоформы // Переработка текста методами инженерной лингвистики. – Минск, 1982.; Муҳаммедов С.А. 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. – Л., 1981.; Айимбетов М.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.; Ризаев С. Ўзбек тилининг лингвостатистик тадқиқи: Фил. фан. док. дисс...автореф. – Тошкент, 2008.; Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.; Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.; Жуманазарова Г.У. Фозил Йўлдош ўғли дostonлари тилининг лингвопоэтикаси: Фил. фан. док. дисс...автореф. – Тошкент, 2017.; Ўринбоева Д.Б. Ўзбек фольклори матнларининг лингвостатистик тадқиқи. – Тошкент: Фан, 2010.; Ўринбаева Д. Халқ оғзаки ижоди: жанрий-лисоний ва лингвостатистик тадқиқ муаммолари: фил. фан. док. дисс. автореф. – Самарқанд, 2019.– 74 б.

¹⁰ Муҳаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. – Тошкент, 2006.

¹¹ Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида). Филол. фан. бўйича фалсафа доктори (PhD)...дис. афтореф. – Тошкент, 2018. – 52 б.

¹² Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2018. – 250 б.

дастурнинг лингвистик модуллари¹³, тил корпуси лингвистик базасини тузиш тамойиллари¹⁴, ўзбек тили атов бирликларини семантик теглашнинг лингвистик асослари¹⁵, ўзбек тили таълимий корпусини тузишнинг лингвистик асослари¹⁶, ўзбек миллий корпусида идиомалар базасини яратиш¹⁷ монографик планда ўрганилган. Шунингдек, ўзбек-инглиз параллел корпуси тузиш муаммолари¹⁸, ўзбек тилидаги отларни автоматик таҳлил қилиш¹⁹, сўз ясалишининг формал моделлари²⁰, корпус лингвистикасида полисемия²¹ ҳақида қатор мақолалар нашр этилган. Ўзбек тилида автоматик морфологик таҳлилни айрим масалалари ёритилган ишлар мавжуд²². Ш.Ҳамроева томонидан эса ўзбек тили морфоанализатори ва унинг лингвистик таъминотини яратиш масаласи тадқиқот объекти сифатида ўрганилган²³.

Бу ишлар компьютер лингвистикаси йўналишидаги бир қатор муаммоларни тадқиқ этганлиги билан аҳамиятли. Аммо ўзбек тили автоматик семантик таҳлил дастури – семантик анализатор ва унинг лингвистик асослари масаласи махсус ўрганилмаган. Диссертацияни ёзиш жараёнида юқорида санаб ўтилган илмий тадқиқотлар пухта ўрганилди ва улардан тадқиқотда фойдаланилди.

Тадқиқотнинг диссертация бажарилган олий таълим ёки илмий-тадқиқот муассасасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги. Диссертация Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёт университетининг “Ўзбек тили имкониятларидан самарали

¹³ Абжалова М.А. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар таҳрири дастури учун): Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Фарғона, 2019. – 164 б.; Абжалова М. Матнларга автолингвистик ишлов бериш тизимлари // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент, 2018. – 320 с.

¹⁴ Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2019. – 140 б.

¹⁵ Ахмедова Д.Б. Атов бирликларини ўзбек тили корпуслари учун лексик - семантик теглашнинг лингвистик асос ва моделлари: Филол. фан. Бўйича фал. доктори (PhD) диссертацияси. – Бухоро, 2020. – 247 б.

¹⁶ Холиёров Ў. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари. Филол. фан. бўйича фал. док-ри. дисс. автореф. – Термиз, 2021. – Б. 16. – 52 б.

¹⁷ Бегматова Г. Ўзбек миллий корпусида идиомалар базасини яратиш. Филол. фан. бўйича фалсафа доктори (PhD) дис.автореф. – Термиз, 2021. – 48 б.

¹⁸ Karimov R., Mengliev B. Theoretical fundamentals of uzbek-english parallel corpus / Journal of critical reviews. ISSN- 2394-5125. – VOL 7, ISSUE 17, 2020. – P. 73-76.; Karimov R.A., Mengliev B.R. The Role of the Parallel Corpus in Linguistics, the Importance and the Possibilities of Interpretation International Journal of engineering and Advanced Technology (IJEAT). ISSN: 2249 – 8958, Volume-8, Issue-5S3 July 2019. – P. 388-391.

¹⁹ Орхун М. Computational analysis of uzbek nouns / Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент, 2018. – 320 с.

²⁰ Турсунов А. Вопросы словообразования в формальных моделях тюркских языков (на примере узбекского языка) // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». Труды конференции) – Ташкент, 2018. – 320 с.

²¹ Абдурахмонова М., Рахманова А. Корпус лингвистикасида полисемия // “Компьютер лингвистикаси: муаммолар, ечим, истиқболлар” Республика I илмий-техникавий конференция. – Vol. 1 №. 01 (2021).

²² Абдурахмонова Н. Automatic morphological analyze for English-Uzbek system // Известия Кыргызский государственный технический университет им. И.Раззакова. Теоретический и прикладной научно-технический журнал. № 2 (38) – Бишкек, 2016. – С. 12-18.

²³ Ҳамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. Филология фан. д-ри. дисс. ав-ти. – Фарғона, 2021. – 76 б.

фойдаланишнинг лингвистик ва дидактик муаммолари” концепциясига мувофиқ бажарилган.

Диссертациянинг мақсади ўзбек тили семантик анализаторининг лингвистик асосларини яратишдан иборат.

Диссертациянинг вазифалари:

лингвистик анализаторнинг умумий тавсифи, турлари ҳамда унинг ахборот-қидирув тизимидаги аҳамиятини ўрганиш;

семантик анализ ва синтезнинг матннинг автоматик таҳлилидаги аҳамиятини ёритиш;

семантик анализаторнинг ахборот-қидирув тизимидаги ўрнини аниқлаш;

ахборот-қидирув тизимида омонимия, полисемия ва полифункционал сўзларни фарқлашнинг лингвистик асосларини аниқлаш;

омонимияни фарқлаш/бартараф этиш масаласини тадқиқ этиш;

полифункционал ва кўп маъноли сўзларни корпусда бартараф этиш учун филтър ишлаб чиқиш;

ўзбек тили семантик анализатори учун омоним, полифункционал ва кўп маъноли сўзлар қуршовини лингвистик моделлаштириш.

Тадқиқот объекти сифатида ўзбек тили семантик анализаторининг лингвистик асосларини ташкил этадиган омоним, полифункционал ва полисемантик сўзлар танланган.

Тадқиқот предметини лингвистик асос яратишда аҳмиятли бўлган омоним, полифункционал ва полисемантик сўзларни лингвистик моделлаштириш ҳамда системалаштириш ташкил этади.

Тадқиқот усуллари. Тадқиқот мавзусини ёритишда тавсифлаш, таснифлаш, қиёслаш, статистик, моделлаштириш методларидан фойдаланилди.

Тадқиқотнинг илмий янгилиги қуйидагилардан иборат:

лингвистик анализаторнинг ахборот-қидирув тизимидаги табиий тилни қайта ишлаш жараёнида намоён бўладиган сўзнинг морфологик, синтактик ва семантик хусусиятларини ифодалашни аниқланиб, семантик анализ ва синтез натижасида матн маълумотлари компьютер тушунадиган лексик-морфологик, синтактик ва семантик ҳодисага айланиши далилланган;

ўзбек тили бирликлари автоматик таҳлилида омонимлик, полифункционаллик ва кўп маънолиликни фарқлашнинг морфологик, синтактик лингвистик асослари яратилиб, от ва феъл орасидаги омонимликда морфологик, сифат ва феъл орасидаги омонимликда морфологик-синтактик, фақат от ва фақат феъл туркуми доирасидаги омонимликда синтактик-морфологик, фақат сифат туркуми доирасидаги омонимликда синтактик, турли туркум доирасидаги омонимликда эса морфологик, синтактик, морфологик-синтактик ва синтактик-морфологик омил етакчилик қилиши аниқланган;

тилдаги омонимлик, полифункционаллик ва кўп маънолиликни тўлиқ бартараф этиш учун фақат соф семантик филтърлар иш бермаслиги, бунда морфо-семантик, синтактик-семантик филтърлар ҳам муҳим аҳамият касб

этиши лингвистик омил асосида аниқланиб, омонимлик, полифункционаллик ва кўп маънолилиқни фарқлаш учун хизмат қиладиган энг зарурий восита – семантик, морфологик, морфо-семантик филтър ишлаб чиқилган;

лингвистик моделлаштиришда сўз семантикаси билан ишлаш анча мураккаб жараён эканлиги, изоҳли луғатда айрим сўзлар маъносининг берилишида чалкашлиқлар ва муаммоларни бартараф этиш мақсадида ўзбек тили семантик анализатори учун омоним, полифункционал ва кўп маъноли сўзлар қуршови дастурий таъминотининг лингвистик жиҳатлари моделлаштирилган;

ўзбек тили семантик анализатори учун омоним сўзлар асосий лингвистик таъминот ҳисобланиб, корпусда омонимлар муаммосини ечиш, омонимияни бартараф этиш учун омоним сўзлар маълумотлар базасини шакллантириш учун тилда мавжуд бўлган омонимларни алифбо тартибида изоҳи билан бериш, морфологик йўл билан фарқланадиган омонимлар учун қўшимчалар комбинациясини ишлаб чиқиш, ўнг ва сўнг бирикувчилар билан фарқланадиган омонимлар учун синтактик филтър базасини яратиш, яратилган филтър базаси асосида омонимларнинг лингвистик моделларини ишлаб чиқиш, лингвистик моделлар орқали математик моделлар яратиш каби семантик анализаторлар юзасидан тавсиялар ишлаб чиқилган.

Тадқиқотнинг амалий натижалари қуйидагилардан иборат:

лингвистик анализаторнинг умумий тавсифи, турлари ҳамда унинг ахборот-қидирув тизимидаги аҳамияти, семантик анализ ва синтезнинг матннинг автоматик таҳлилидаги ўрни далилланган;

семантик анализаторнинг ахборот-қидирув тизимидаги ўрни, жумладан, сўзнинг тўғри маъносини танлаш ва танланган маъноларни мустақил семантик изоҳга эга бўлган сўзларга боғлаш каби вазифалар аниқланган;

ахборот-қидирув тизимида омонимия, полисемия ва полифункционал сўзларни фарқлашнинг морфологик, синтактик омиллари аниқланиб, лингвистик асослари яратилган;

корпусда омонимия, полифункционал ва кўп маъноли сўзларни фарқлаш/бартараф этиш учун морфологик, морфо-семантик, морфо-синтактик филтърлар ишлаб чиқилган.

семантик анализатор учун омоним, полифункционал ва полисемантик сўзлар қуршови дастурлаш учун лингвистик моделлаштирилган.

Тадқиқот натижаларининг ишончлилиги ўрганилган материалларнинг ўзбек тили табиатидан келиб чиққан ҳолда хулосалар қилишга ёрдам берганлиги, уларнинг асосли эканлиги, методологик мукамаллиги, ўзбек тили семантик анализаторининг лингвистик асосларини ишлаб чиқишда амалда исботланган манбаларга таянилганлиги билан изоҳланади.

Тадқиқот натижаларининг илмий ва амалий аҳамияти. Тадқиқот ўзбек тили семантик анализатори яратишнинг назарий асослари ва унинг лингвистик асосларини ишлаб чиқишда, компьютер лингвистикаси йўналишида тадқиқотлар яратишда назарий аҳамият касб этади.

Тадқиқотнинг амалий аҳамияти компьютер лингвистикаси, корпус лингвистикаси ҳамда амалий тилшунослик фанлари ўқитилиш жараёнида дастур, режалар тузиш ҳамда мавзуларни баён этишда манба вазифасини ўташи, ўзбек тили семантик анализаторини ишлаб чиқишда, ўзбек тилида турли типдаги корпусларни семантик теглашда қўлланилиши мумкинлиги билан изоҳланади.

Тадқиқот натижаларининг жорий қилиниши.

Семантик анализаторнинг ахборот-қидирув тизимидаги аҳамияти, ахборот-қидирув тизимида омоним, полифункционал ҳамда кўп маъноли сўзларни фарқлашнинг лингвистик асослари, ўзбек тили семантик анализатори учун омоним, полифункционал ва кўп маъноли сўзлар қуршовини лингвистик моделлаштириш масалалари натижаларидан:

ўзбек тили семантик анализатори учун омоним сўзлар асосий лингвистик таъминот ҳисобланиб, корпусда омонимлар муаммосини ечиш, омонимияни бартараф этиш учун омоним сўзлар маълумотлар базасини шакллантириш муҳим масала бўлиб, бунинг учун тилда мавжуд бўлган омонимларни алифбо тартибида изоҳи билан бериш (1); морфологик йўл билан фарқланадиган омонимлар учун қўшимчалар комбинациясини ишлаб чиқиш (2); ўнг ва сўл бирикувчилар билан фарқланадиган омонимлар учун синтактик филтър базасини яратиш (3); яратилган филтър базаси асосида омонимларнинг лингвистик моделларини ишлаб чиқиш (4); лингвистик моделлар орқали математик моделлар яратиш (5) каби семантик анализаторлар юзасидан тавсиялардан Ўзбекистон Республикаси Фанлар Академияси Қорақалпоғистон бўлими Қорақалпоқ гуманитар фанлар илмий-тадқиқот институти томонидан олиб борилган ФА-Ф1-Г003 “Ҳозирги қорақалпоқ тилида функционал сўз ясалиши” мавзусидаги фундаментал илмий лойиҳада фойдаланилган (Ўзбекистон Республикаси Фанлар Академияси Қорақалпоғистон бўлимининг 2021 йил 17 октябрдаги 17.01/265-сон маълумотномаси). Натижада ҳозирги қорақалпоқ тилида сўз ясалиш моделлари мукамаллаштирилган; қорақалпоқ тилидаги омоним, полисемия ва полифункционал сўз маъноларини автоматик фарқлаш тамойиллари ишлаб чиқилган;

лингвистик анализаторнинг ахборот-қидирув тизимида табиий тилни қайта ишлаш жараёнида намоён бўладиган сўзнинг морфологик, синтактик ва семантик хусусиятларини ифодалашни аниқланиб, семантик анализ ва синтез натижасида матн маълумотлари компьютер тушунадиган лексик-морфологик, синтактик ва семантик ҳодисага айланишига доир хулосалардан Ўзбекистон Республикаси Фанлар Академияси Қорақалпоғистон бўлими Қорақалпоқ гуманитар фанлар илмий-тадқиқот институти томонидан олиб борилган ФА-А1-Г007 “Қорақалпоқ нақл-мақоллари лингвистик тадқиқот объекти сифатида” мавзусидаги илмий-амалий лойиҳада фойдаланилган (Ўзбекистон Республикаси Фанлар Академияси Қорақалпоғистон бўлимининг 2021 йил 17 октябрдаги 17.01/265-сон маълумотномаси). Натижада Қорақалпоқ нақл-мақолларининг лингвистик моҳиятини

аниқлашда ахборот-қидирув тизимида бу бирликларни кенг қўллаш тамойиллари ишлаб чиқилган;

ўзбек тили бирликлари автоматик таҳлилида омонимлик, полифункционаллик ва кўп маънолиликни фарқлашнинг морфологик, синтактик лингвистик асослари яратилиб, от ва феъл орасидаги омонимликда морфологик, сифат ва феъл орасидаги омонимликда морфологик-синтактик, фақат от ва фақат феъл туркуми доирасидаги омонимликда синтактик-морфологик, фақат сифат туркуми доирасидаги омонимликда синтактик, турли туркум доирасидаги омонимликда эса морфологик, синтактик, морфологик-синтактик ва синтактик-морфологик омил етакчилик қилиши хусусидаги натижаларидан Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетида бажарилган I-ОТ-2019-42 рақамли “Ўзбек ва инглиз тилларининг электрон (Инсон қиёфаси, феъл-атвори, табиат ва миллий тимсоллар тасвири) поэтик луғатини яратиш” мавзусидаги фундаментал илмий лойиҳада фойдаланилган (Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетининг 2021 йил 23 октябрдаги 04/1-1995-сон маълумотномаси). Натижада инсон қиёфаси, феъл-атвори, табиат ва миллий тимсоллар тасвирида омоним, полифункционал ва кўп маъноли сўзлар ўрни ва ролини ёритишга, ўзбек ва инглиз тилларидаги омоним, полифункционал ва кўп маъноли сўзларнинг семантик ва структур хусусиятларини очиқ беришга эришилган;

тилдаги омонимлик, полифункционаллик ва кўп маънолиликни тўлиқ бартараф этиш учун фақат соф семантик филтрлар иш бермаслиги, бунинг учун морфо-семантик, синтактик-семантик филтрлар ҳам муҳим аҳамият касб этиши лингвистик омил асосида аниқланиб, омонимлик, полифункционаллик ва кўп маънолиликни фарқлаш учун хизмат қиладиган энг зарурий восита – семантик, морфологик, морфо-семантик филтр хусусидаги хулосалардан 5220100 – Филология ва тилларни ўқитиш (ўзбек тили) бакалавр таълим йўналиши талабаларига мўлжалланган “Ҳозирги ўзбек тили” дарслигининг “Лексика, лексик-семантик сатҳ ва лексикология” деб номланган 11-параграфи ҳамда “Ўзбек тили лексикаси таснифи” деб номланган 18-параграфини ёзишда фойдаланилган (Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетининг 2021 йил 20 октябрдаги 04/1-1975-сон маълумотномаси). Натижада ўзбек луғатчилигига доир фикрларни билдиришда корпусда йиғилган омоним, кўп маъноли ва полифункционал сўзлар ҳамда уларнинг филтри, лингвистик модели нафақат лексикология, балки семантик анализаторнинг лингвистик асослари тилшуносликнинг морфология ва синтаксис сингари йўналишларига бой илмий манба бўла олиши хусусидаги хулосалар мукамаллашган;

лингвистик моделлаштиришда сўз семантикаси билан ишлаш анча мураккаб жараён эканлиги, изоҳли луғатда айрим сўзлар маъносининг берилишида чалкашликлар ва муаммоларни бартараф этиш мақсадида ўзбек тили семантик анализатори учун омоним, полифункционал ва кўп маъноли сўзлар қуршови дастурий таъминот алгоритми лингвистик жиҳатлари

ҳақидаги хулосалардан Бухоро вилояти телерадиокомпанияси “Бухоро” телеканалининг “Ассалом, Бухоро!”, “Саккизинчи мўъжиза”, “Адабий муҳит” кўрсатувларини тайёрлашда фойдаланилган (Бухоро вилояти телерадиокомпаниясининг 2021 йил 15 октябрдаги № 1/267-сон маълумотномаси). Натижада кўрсатувларнинг илмий асосланганлиги, оммабоплиги ва томошабинбоплигини таъминлашда муҳим аҳамият касб этиб, тил маданиятини шакллантириш ҳамда замонавий ахборот тизимида ўзбек тилининг имкониятларидан фойдаланиш ҳақидаги хулосалар мукаммаллашган.

Тадқиқот натижаларининг апробацияси. Мазкур тадқиқот натижалари 15 та халқаро, 10 та республика анжуманида муҳокамадан ўтказилган ҳамда куйидаги профилларда ҳаволага эришилган: <https://orcid.org/0000-0002-2045-1912>, <https://www.researchgate.net/profile/Shahnoza-Gulyamova>, <https://scholar.google.com/citations>.

Тадқиқот натижаларининг эълон қилиниши. Диссертация мавзуси бўйича 49 та илмий иш нашр этилган, жумладан, 2 та монография, Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясининг докторлик диссертациялари асосий илмий натижаларини чоп этиш учун тавсия этилган илмий нашрларда 17 та илмий мақола, улардан 5 таси хорижий журналларда чоп қилинган.

Диссертациянинг ҳажми ва тузилиши. Диссертация кириш, тўрт боб, хулоса, фойдаланилган адабиётлар рўйхати ва иловадан иборат. Ҳажми 236 саҳифани ташкил этади.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Кириш қисмида мавзунинг долзарблиги ва зарурати асосланган, тадқиқотнинг республика фан ва технологиялари ривожланишининг устувор йўналишларига боғлиқлиги кўрсатилган, мақсад ва вазифалари берилган, объекти ва предмети тавсифланган, илмий янгилиги ва амалий натижалари баён қилинган, натижаларнинг илмий ва амалий аҳамияти очиқ берилган, жорийланиши, апробацияси, нашр этилган ишлар ва диссертация тузилиши бўйича маълумотлар келтирилган.

Диссертациянинг биринчи боби “**Семантик анализаторга доир назарий масалалар**” деб номланган. Бобнинг “Лингвистик анализатор тушунчаси, унинг турлари ва тавсифи” номли биринчи бўлимида лингвистик анализаторнинг умумий масалалари, лингвистик анализатор турлари, дунё ва ўзбек компьютер тилшунослигида анализатор тадқиқи масалалари таҳлил этилган. Фойдаланувчилар интернет тармоқларидаги маълумотдан турли мақсадда фойдаланишади. Одатда, ахборот объекти деб аталадиган аниқ маълумот керак бўлади. Бу объектлар турли типларга бўлиниб, айрим тоифаларидаги фойдаланувчиларда қизиқиш уйғотади. Табиий тилдаги матнлар оқимида бу объектларни қидириш кўп вақтни талаб этади, бу эса аниқ дастурий муаммоларни ҳал қилишни секинлаштиради. Матнларни бир

тилдан иккинчи тилга таржима қилиш, турли тилларни деконвертлаш, компьютер дарсликлари, белгили матн корпорацияси, матнларни анализ ва синтез қилиш, аннотация ва абстракциялаш дастурларини яратиш зарурати табиий тилга автоматик ишлов бериш муаммосини бартараф этиш вазифасини кўяди.

Лингвистик анализатор – автоматлашган қидирув тизимида табиий тилни қайта ишлаш жараёнида сўзнинг морфологик, синтактик ва семантик хусусиятини ифодалайдиган восита. Анализатор тушунчаси термин сифатида кенг маънода қўлланади. Жумладан, лингвистика билан бир қаторда биология, физиология, анатомия, психология, физика, кимё каби ўнлаб соҳаларда маълум мақсад ва маънода ишлатилади. Лингвистикада эса анализатор табиий тилни автоматик қайта ишлашда муҳим ўрин тутди. Табиий тилни қайта ишлаш (Natural Language Processing (NLP)) – бу сунъий интеллект ва математик тилшуносликнинг умумий йўналиши бўлиб, табиий тиллардаги матнларни компьютерда анализ ва синтез қилиш муаммоларини ўрганади. Ушбу муаммони ҳал қилиш инсон ва компьютер ўртасида ўзаро алоқанинг янада қулай шаклини яратишни англатади. Лингвистик анализаторни қуйидаги турларга бўлиб ўрганиш мақсадга мувофиқ: 1) семантик анализатор; 2) морфологик анализатор; 3) синтактик анализатор.

Катта гуруҳлар ўз ичида лемматизатор, морфоанализатор, стеммер, парсер, орфокорректорларни олади. Бу анализаторларнинг ҳар бири ишлаш принципи, мақсади, вазифасига эга.

Бобнинг “Семантик анализ ва синтезнинг матн автоматик таҳлилидаги аҳамияти” деб номланган иккинчи бўлимида ахборот-қидирув тизимида матннинг лингвистик таҳлили, автоматик таҳлилда семантик анализ ва синтез масаласи ҳамда АҚТда семантик қидирув ва унинг муаммолари каби масалалар ҳақида фикр юритилган. Сунъий равишда сўровни шакллантириш, тушуниш мақсадида табиий тилда гапнинг таҳлили, синтези (SQL, SPARQL, қидирув сўровлар ва бошқалар), гапнинг бир табиий тилдан бошқа табиий тилга таржима қилиниши (машина таржима тизими), матннинг семантик таҳлили вазифалари – TextMining, матн ресурсларининг семантик классификацияси каби масалалар амалий аҳамиятга моликдир.

И.В.Смирнов семантика моделлари сифатида қуйидагиларни кўрсатади: 1) лексик семантика – сўзлар талқини (изоҳи); 2) дистрибутив семантика – сўзларнинг бошқа сўзлар билан бирикиши орқали ифодалайдиган маъноси; 3) “маъно-матн” модели; 4) ифодалар семантикаси; 5) семантик фреймлар модели²⁴.

Кўрсатилган семантика моделлари семантик анализаторнинг фаолияти учун муҳим ҳисобланади. Чунки семантик анализатор – сўз маъноси билан иш кўрадиган восита. Бизнингча, ўзбек тили семантик анализатори ҳам бевосита юқоридаги ҳолатларни ҳисобга олган ҳолда иш юритади, семантик

²⁴Смирнов И.В. Обработка естественного языка и искусственный интеллект // <https://cs.hse.ru/data/2017/12/06/1161306556/Slides-IntroToAI-HSE-2017-09-Smirnov.pdf>

таҳлилни амалга оширади. Чунки ушбу хусусиятлар бирлашиб, семантик белгини аниқ кўрсатишга хизмат қилади.

Матннинг автоматик таҳлили бир қатор мураккаб операциялардан иборат. Матнни автоматик таҳлил қилиш натижасида унинг маълумотлари компьютер тушунадиган лексем-морфологик, синтактик ва семантик тасаввурга айланади. Автоматик матнни таҳлил ва синтез қилиш муҳим аҳамиятга эга. Чунки компьютер лингвистикасининг ҳам назарий, ҳам амалий тараққиёти инсон эҳтиёжларини амалга оширишга қаратилган самарали машина таржима тизимлари яратишга замин яратади.

Семантик таҳлил автоматик матн таҳлилининг энг мураккаб йўналишидир. Сўзлар ўртасида матнда турли тил ифодаларини бирлаштириш учун семантик алоқаларни ўрнатиш керак.

Семантик таҳлил асосида сўзнинг маъноси элементар семантик бирлик эмас, деган тасдиқ ётади. У кўпроқ элементар маъноларга – семантик тил луғатининг бирликларига бўлинади. Семантик тилнинг бу бирликлари турли комбинациялардан ўзига хос атомлар бўлиб, табиий тилнинг реал сўзларидан “молекулалар” ҳосил қилади²⁵.

Айнан семантик таҳлил барча тил сатҳларини автоматик таҳлил қилишда вужудга келадиган омонимия муаммосини ҳал этади:

1. Лексик омонимия: умумий маъно элементларига эга бўлмаган сўзларда товуш ва ёки имлонинг тўғри келиши.

2. Морфологик омонимия: қўшимчалар омонимлиги.

3. Лексик-морфологик омонимия (омонимиянинг энг кўп тарқалган тури): икки хил лексема сўз шаклининг тўғри келиши.

4. Синтактик омонимия: кўп маъноликка эга бўлган бир синтактик тузилма²⁶.

Семантик таҳлил модули қуйидаги икки синфдан иборат: DictionaryOr ва Semantic-Analysis. DictionaryOr семантик таҳлил луғати билан ўзаро алоқа қилиш учун масъул бўлган статистик синф ҳисобланади. Semantic-Analysis (семантик таҳлил) қабул қилинган гапнинг таҳлилини амалга оширади, ундан кейин семантик тармоқни қайтаради.

Семантик синтез – бир иборанинг семантик жихатдан синтактик тузилишга ўтиши (синтактик структурадан иборани ифодаловчи сўз шакллариининг лексик ва грамматик хусусиятлари занжирига ўтиши; лексик-морфологик тавсифдан ҳақиқий сўз шаклига ўтиши)²⁷.

Қидирув тизимлари СК (семантик қидирув)ни амалга оширишда жуда кўп муаммоларга дуч келади. Жумладан, қидирув сўровини киритишда фойдаланувчи айнан нимани назарда тутганини, яъни турли контекстларда сўз ёки иборанинг бир қанча маънолари мавжудлигини аниқлаш муаммоси. Семантик қидирув компьютер йўли билан ҳал қилиб бўлмайдиган муаммоларни ечишда ёрдам беролмайди. Семантик қидирув нафақат

²⁵ Апресян Ю.Д. Идеи и методы современной структурной лингвистики. М.: Просвещение, 1966. – С.254.

²⁶ Большакова Б.И. Компьютерная лингвистика: методы, ресурсы, приложения // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. М.: МИЭМ, 2011. С. 93—94.

²⁷ Щипицина Л.Ю. Информационные технологии в лингвистике: учеб. пособие / Л.Ю.Щипицина. – М.: ФЛИНТА: Наука, 2013. – 128 с. – С. 27.

контекстни, балки бошқа омиллар (жумладан, жорий вазият, фойдаланувчи жойи, семантик кидирувда сўзларнинг ўзгариши, синонимлар, умумий ва ихтисослаштирилган сўровлар, сўровлар учун “оддий тил”) таҳлилини ҳам инобатга олади.

Бобнинг “Семантик анализатор ва унинг ахборот-кидирув тизимида тутган ўрни” номли учинчи бўлимида семантик анализаторнинг ўрганилиши, мақсад ва вазифалари, синтактик-семантик анализаторлар, ўзбек тили семантик анализаторининг лингвистик таъминоти масаласи тавсифланган.

Семантик анализатор – сўровнинг синтактик тузилишини концептуал графиклар ёрдамида семантик жиҳатдан қайта ишлаш дастури. Концептуал график семантик ёки бошқача қилиб айтганда, табиий тилни тушуниш моделларида вазият ва билимни концептуал тасаввур қилишдир. Графиканинг тугунлари категория ва тушунчани ифодаловчи лексик birlikлар бўлиб, бир-бирига формал ва ассоциатив муносабат орқали боғланади. Графикдаги муносабатларнинг йўналиши ҳар доим юқоридан пастга – умумлаштиришнинг юқори даражадаги тушунчаларидан уларни тавсифловчи тушунчаларга йўналтирилади. Элементар маъно графикнинг бир-бирига боғланган қўшни тугунлари сифатида аниқланади. Бундай боғланишлар ҳар доим ҳам номланмайди, баъзан улар фақат икки сўз (талаба-институт, қайин-дарахт) ўртасидаги ўзаро таъсирнинг ҳосиласини қайд этади. Графа маълум бир сўзнинг лексик фаол ва пассив валентликларини очиб берувчи, рухсат этилган бирикмаларга кирувчи ўзаро боғланган элементар маънолар мажмуасидан иборат. Графанинг икки тугунни бирлаштирувчи боғланган қисми кичик гуруҳни ташкил қилади. Семантик анализатор материални қайта ишлаш жараёнининг охириги босқичидир. Лингвистик процессор тугагандан сўнг таҳлил қилинган ахборот индексация босқичига ўтказилади.

Семантик анализатор сўз маъносининг тўғрилигини текширадиган тизим экан, уни бевосита “тафаккур анализатори” деб аташ мумкин. Бизнингча, семантик анализатор қуйидаги имкониятларга эга бўлиши керак:

биринчидан, тўғри ёзишнинг (орфография) жуда кучли текширув воситаси;

иккинчидан, сўз маъносини аниқлаштирувчи тизим;

учинчидан, сўзларни, уларнинг маъноси бироз бузилса ҳам, асосли конструкцияларга йиғишга ҳаракат қилади;

тўртинчидан, айрим номаълум сўзларни ёки нотўғри жойлаштирилган вергулларни ўз ичига олган структурани гапдан чиқаришга ҳаракат қилади.

Кўринадики, семантик анализатор яратиш анча мураккаб вазифа. Бунинг учун биз даставвал семантик анализатор учун лингвистик таъминот масаласини ҳал қилишимиз зарур.

Семантик анализатор – бу “таҳлилнинг юраги”: у дастурлаш тили билан боғлиқ барча қоидаларни билади. Семантик анализатор тилда мавжуд бўлган омоним, полифункционал ва кўп маъноли сўзларнинг маъносини фарқлашга хизмат қиладиган восита.

Диссертациянинг II боби “**Ахборот-қидирув тизимида омонимликни фарқлашнинг лингвистик асослари**” деб номланган. Бобнинг “Анъанавий тилшунослик ва компьютер лингвистикасида омонимия тадқиқи” номли биринчи бўлимида жаҳон тилшунослигида, туркологияда ҳамда ўзбек тилшунослигида омонимиянинг ўрганилиши ҳақида фикр юритилган.

С.Ульманн омонимияни “лингвистик универсал” деб атаган. У омонимия мутлақ универсал эмаслиги, балки жаҳонда омонимия мавжуд бўлмаган тилни ҳам топиш мумкинлиги, омонимсиз “бундай тил янада самарали алоқа воситаси” бўлишини таъкидлайди. С.Ульманнинг фикрича, омонимия статистик универсалия, яъни нисбий. Шунингдек, олим дунёнинг кўпгина тилларида омонимлар мавжуд, деб ҳисоблаган [Ульманн. 1970].

Жаҳон тилшунослигида омонимия ҳодисасининг ўрганилиши борасидаги назарияларни қуйидагича гуруҳлаш мумкин: 1. Омонимия ҳодисаси тилда мавжуд бўлган ҳамда аҳамиятга эга бўлган воқелик. 2. Тил учун аҳамиятга эга бўлмаган, “чалғитувчи” ҳодиса.

Омонимия К.Аханов (қозоқ тили) [1956], М.Миртожиев (ўзбек тили) [1964], А.Х.Ахтямов (бошқирд тили) [1966], Р.Бекджанова (қирғиз тили) [1967], Б.С.Салимгареева (татар тили) [1972], А.Гельдимурадов (туркман тили) [1974], А.Меметов (қрим-татар тили) [1977], А.А.Гасанов (озарбайжон тили) [1980] каби турколог олимларнинг тадқиқоти мавзуси бўлди. Бу турколог олимлар А.И.Смирницкий, В.В.Виноградов, О.С.Ахманова ва бошқа тилшунос олимлар ишлаб чиққан омонимия назариясига таянадилар. Булардан ташқари, туркологияда омонимлар муаммоси Н.А.Баскаков, Н.К.Дмитрийнинг илмий асарларида ҳам таҳлил этилган. Шунингдек, В.В.Радлов, Э.В.Севортян, А.А.Юлдашева, К.К.Юдахина, М.Х.Ахтямова, С.И.Ибрагимов каби қатор рус ва ўзбек олимлари томонидан тадқиқотлар олиб борилган.

Ҳар бир тилда омонимлар лексикологиянинг тарихий ҳодисаси бўлиб, улар тилнинг қашшоқлиги кўрсаткичи эмас, балки лексик захиранинг бойлиги ва хилма-хиллиги кўрсаткичидир.

Бобнинг “Корпусда омонимияни фарқлаш/бартараф этиш масаласи” деб номланган иккинчи бўлимида омонимларнинг компьютер лингвистикасидаги тадқиқи ҳамда уларни семантик таҳлил қилиш тажрибаси, омонимларни таҳлил қилиш усуллари, ўзбек тили семантик анализатори учун омоним сўзлар маълумотлар базасини шакллантириш масаласи таҳлил қилинган. Маълумки, тил ҳодисаларини ўрганиш, тадқиқ ва таҳлил этиш ҳамма даврда ўз аҳамиятига эга. Жаҳон тилшунослигида омонимия, синонимия, антонимия ва полисемия каби ҳодисалар ўзбек тилшунослиги учун янги ҳисобланган компьютер тилшунослиги нуқтаи назаридан кенг доирада ўрганилган, унинг амалий натижаларидан корпус тузиш ва такомиллаштириш каби мақсадларда фойдаланилмоқда. Тилшуносликнинг илк ва ўрта босқичларида омонимия ҳодисасининг пайдо бўлиши, тараққиёти ва ёндош ҳодисаларга муносабатлари ўрганилган бўлса, замонавий лингвистикада унинг корпусда берилиши масаласи кун тартибига

қўйилди. Бу вазифани бажариш учун мавжуд луғатлар ахборот банки вазифасини ўтайди.

Корпус лингвистикасида омонимлик муаммосини ечиш, омоним бирликларни теглаш ва матнни автоматик ўқиш жараёнида омонимияни бартараф этиш масаласи (“снятие омонимии”)га оид қатор тадқиқотлар вужудга келган. Жумладан, Г.И.Кустова, О.Н.Ляшевская, Е.В.Падучева, Е.В.Рахилина, Б.П.Кобрицов, Т.И.Резникова²⁸, В.В.Куканова²⁹, А.А.Кретов³⁰лар шу масалалар ечимига бағишланган қатор ишларни эълон қилишган.

Омонимияни бартараф этиш компьютер тилшунослигининг кўплаб дастурларида, хусусан, қидирув тизимларида аҳамиятли. Чунки у муайян сўров синфларини қайта ишлаш аниқлигини ошириши ёки сақланган ахборот миқдорини камайтириши мумкин. Инглиз тилидаги матнларда омонимияни бартараф этишни қиёслаш мақсадида икки метод танлаб олинган: таянч вектор методи ва яширин Марков модели (ЯММ). Бу методлар луғатда бўлмаган сўзлар учун омонимияни қисман бартараф этиш имкониятини беради. Таниқли яширин Марков модели (Hidden Markov Model (HMM) tagging) омонимларни эҳтимолий ёндашув асосида аниқлашга қаратилган. Алгоритмнинг асосий ғояси шундан иборатки, жумлага киритилган ҳар бир сўз учун қуйидаги функцияни максималлаштирадиган грамматик синфни (тегни) танлаш керак:

$$P(\text{word/tag}) * P(\text{tag/previous } n \text{ tags})$$

$P(\text{tag/previous } n \text{ tags})$ – шартли эҳтимоллик, ушбу тег олдинги n тег аниқлангандан кейин ҳосил бўлади.

$P(\text{word/tag})$ – шартли эҳтимоллик, берилган жойда ҳосил бўлган $word$ сўзи, берилган грамматик синф тегига эга.

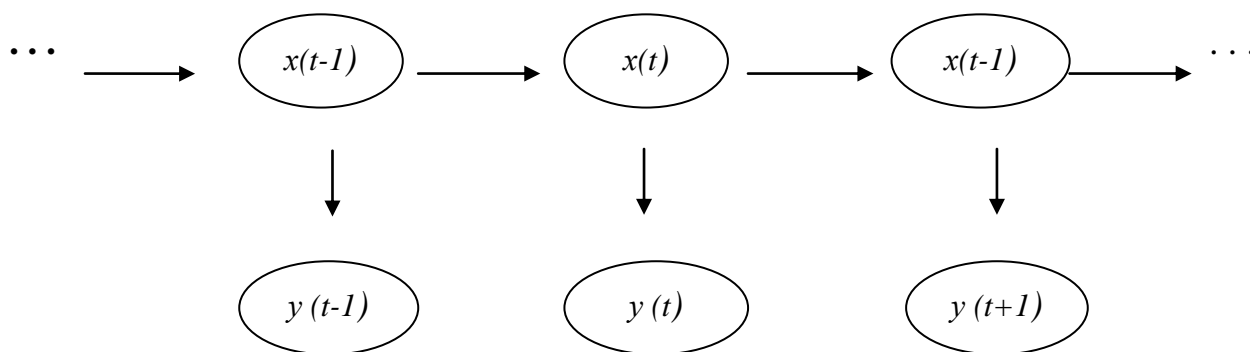
Яширин Марков модели – номаълум параметрларга эга бўлган жараённинг ишлашини симуляция қилувчи статистик модел. Номаълум параметрларни ўзгарувчан кузатувлар асосида ҳал қилиш вазифа қилиб қўйилади³¹.

²⁸Кустова Г.И., Ляшевская О.Н., Падучева Е.В.,Рахилина Е.В. Семантическая разметка лексик в Национальном корпусе русского языка: принципы, проблемы, перспективы// Национальный корпус русского языка:2003-2005. Результаты и перспективы. М.:Индрик, 2005. – С. 155-174.; Ляшевская О.Н. Топологические классы именв семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). СПб.: С.-Петербургский гос. университет, Факультет филологии и искусств, 2008. – С. 276-284.; Рахилина Е.В., Кобрицов Б.П., Кустова Г.И., Ляшевская О.Н., Шеманаева О.Ю. Многозначность как прикладная проблема: семантическая разметка в Национальном корпусе русского языка // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог-2006». – М., 2006. – С. 445-450.; Рахилина Е.В., Кустова Г.И., Ляшевская О.Н., Резникова Т.И., Шеманова О.Ю. Задачи и принципы семантической разметки лексик в НКРЯ // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы/ отв. ред. В.А.Плунгян. СПб.: Не-стор-История, 2009. – С.215-239.; Ляшевская О.Н. Топологические классы именв семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). СПб.: С.-Петербургский гос. университет, Факультет филологии и искусств, 2008. – С. 276-284.; Кобрицов Б.П. Модели многозначности русской предметной лексики: глобальные и локальные правила разрешения омонимии. Автореф. канд. филол. наук.-М.: РГГУ, 2004.; Кобрицов Б.П., Ляшевская О.Н., Шеманаева О.Ю. Поверхностные фильтры для разрешения семантической омонимии в текстовом корпусе // Кобозева И.М., Нариньяни А.С., Селегей В.П. (ред.), Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции Диалог'2005.-М., 2005.

²⁹Куканова В.В. Принципы семантической разметки национального корпуса калмыцкого языка // http://kalmcorp.ru/sites/default/files/kukanova_25.pdf

³⁰Кретов А.А. Анализ семантических помет в НКРЯ // <http://ruscorpora.ru/sbornik2008/11.pdf>

³¹Ветров Д. Скрытые марковские модели // http://www.machinelearning.ru/wikiimages/8/83/GM12_3.pdf.



1-расм. ЯММнинг умумий структураси.

Бунда $x(t)$ яширин ўзгарувчининг қиймати фақат $x(t-1)$ яширин ўзгарувчининг қийматига боғлиқ эканлиги равшан бўлади. Бу боғлиқлик (тобелик) “Марков хусусияти” деб аталади. Шу билан бирга, кузатилган ўзгарувчининг қиймати $y(t)$ фақат яширин ўзгарувчининг $x(t)$ қийматига боғлиқ бўлади. Луғатда бўлмаган сўзларни таҳлил қилиш учун уч ҳарfli постфикслар ишлатилади. Натижада, морфологик хусусиятларни аниқлаш муаммосини ҳал қилиш яширин қийматларнинг энг муносиб қийматлари кетма-кетлигини ҳисоблаш учун камаяди:

$$y'(t) = \operatorname{argmax} P(y/x)$$

ЯММ алгоритми жуда юқори ҳисоблаш мураккаблигига эга (классик Витерби динамик дастурлаш усули билан амалга оширилади, шунга ўхшаш Левенштейн масофани ҳисоблаш учун ишлатилади) ва амалда, одатда, уни камайтириш учун турли соддалаштирувчи тахминлар ишлатилади (масалан, уч сўзли кетма-кетликлардан ортиқ бўлмаслиги керак).

Бобнинг “Омонимликни фарқлаш омиллари” деб номланган учинчи бўлимида омонимликни фарқлашнинг морфологик омиллари, коллокация ҳодисаси ва омонимия, омонимликни фарқлашнинг синтактик омиллари хусусида фикр юритилади. Маълумки, тил бирликлари бир-биридан қатор омиллар билан фарқланади. Жумладан, лексик, морфологик, синтактик факторлар шулар сирасига киради. Бирликларни фарқлашда ҳар бир омилнинг ўз ўрни ва аҳамияти бўлади. Уларни фарқлашда қайсидир омил етакчилик қилса, бошқа омил кучсизланиши мумкин ёки омиллар ҳамкорлиги асосида фарқланиши кузатилади. Тадқиқот объекти бўлган омонимларни фарқлаш бир қатор йўллар орқали амалга ошади. Олдин таъкидланганидек, семантик анализаторнинг омонимларни фарқлаши учун дастлаб уларнинг лингвистик фильтри ва моделлари ишлаб чиқилиши лозим. Бу эса лингвистик билимни, ёндашувни талаб этади. Зотан, табиий тилни қайта ишлаш (NLP) сунъий ақл синтези ва математик тилшуносликнинг энг муҳим ва долзарб йўналишларидан биридир. Бу ўринда асосий ҳаракат инсоннинг табиий тилида ифодаланган матнни таҳлил қилиши ва компьютернинг ушбу маълумотни “идрок этиш”, қайта ишлаш қобилиятини топиш ҳамда такомиллаштиришга қаратилган.

Табий тилни қайта ишлаш – бу сунъий интеллект билан боғлиқ вазифа, чунки у атрофдаги дунё ва инсон ҳиссиётларини юқори даражада тушунишни ўз ичига олади. Шубҳасиз, ушбу муаммоларни ҳал қилиш тадқиқотчидан тилшунослик ҳамда сунъий интеллект соҳасида етарлича билимни талаб қилади.

Тадқиқот учун Ш.Раҳматуллаевнинг “Омонимларнинг изоҳли луғати” ҳамда “Ўзбек тилининг изоҳли луғати” (5 жилдли)да мавжуд омонимлар лингвистик база бўлиб хизмат қилди. Уларни туркумига асосланиб, қуйидагича гуруҳлаб чиқдик: 1) бир сўз туркуми доирасидаги омонимлар; 2) от // феъл ҳамда сифат // феъл туркуми орасида кузатиладиган омонимлар; 3) турли туркум доирасида вужудга келадиган омонимлар.

От // феъл ҳамда сифат // феъл туркуми доирасида учрайдиган омоним сўзлар, асосан, морфологик омил асосида фарқланади. МО бир қатор морфологик воситалар орқали юзага чиқади.

Омоним сўзни фарқлашда морфологик ва синтактик омиллар муҳим ўрин тутади. От ва феъл орасидаги омонимликда морфологик; сифат ва феъл орасидаги омонимликда морфологик-синтактик; фақат от ва фақат феъл туркуми доирасидаги омонимликда синтактик-морфологик; фақат сифат туркуми доирасидаги омонимликда синтактик; турли туркум доирасидаги омонимликда эса морфологик, синтактик, морфологик-синтактик ва синтактик-морфологик омил етакчилик қилиши аниқланди.

Тадқиқотнинг III боби **“Ахборот-кидирув тизимида полифункционал ҳамда кўп маъноли сўзларни фарқлашнинг лингвистик асослари”** деб аталади. Биринчи бўлим “Анъанавий тилшуносликда полифункционал ва кўп маъноли сўзлар тадқиқи” деб номланади.

Полисемия ҳар қандай тилда салмоқли ўрин тутади. Тил бойлиги фақат ибора, мақоллар билан эмас, балки сўзларнинг лексик (луғавий) маънолари билан ҳам ўлчанади. Сўз маъноларининг ранг-баранглиги, яъни сўзларнинг кўп маънолилиги – полисемия тил бойлигида ўз ўрнига эга.

Тилдаги полисемия ҳодисаси жуда қадимдан олимлар диққат марказида бўлган. Машҳур лексикограф Маҳмуд Кошғарий “Девону луғатит турк” китобида туркий полисемантик сўзларни қайд этади. Буюк мутафаккир Алишер Навоий ҳам “Муҳокамат ул-луғатайн” асарида сўз маъноларига алоҳида тўхталади. Навоий асарларида деярли барча сўзлар кўп маъноли, шунинг учун ҳам лексикографлар сўз маъносини изоҳлашда доим полисемантизмга дуч келади. Натижада, ўзбек луғатчилиги тарихида сўз маъносини изоҳлаш ўзининг изчил бир тизимига эга бўлди. Бу даврда яратилган барча луғатларда сўз маъносини изоҳлаш “умумийликдан хусусийликка” тамойилига таянган ҳолда олиб борилган: олдин умумий, тўғри маъно кўрсатилади, сўнгра сўзнинг кўчма маънолари изоҳланади³².

Полисемияни баҳолаш тилшунослар орасида зиддиятли фикрларга сабаб бўлди. Баъзи олимларнинг фикрича, идеал тилда сўз фақат битта маънога эга

³²Нурмонов А. Ўзбек тилшунослиги тарихи. – Тошкент: Ўзбекистон, 2002. – Б. 48. – 232 б.

бўлиши ва ҳар бир маъно алоҳида номга эга бўлиши керак. Бироқ бу бир қарашда қулай кўриниши мумкин, аслида сўзларнинг бирёқламалиги тил имкониятларини пасайтириб, уни миллий ўзликдан маҳрум этади. Кўпчилик олимлар ҳақли равишда сўзларнинг кўп маънолилигини тилнинг заифлиги эмас, балки тил бойлигининг бир кўриниши деб биладилар. Кўп маънолилик тилнинг чексиз имкониятларини эътироф қилади, чунки сўз бойлиги нафақат сўзлар, балки уларнинг қийматлари хилма-хиллигида, сўзларнинг янги семантик фарқларни олиш қобилиятидир. Сўздаги янги маъноларнинг ривожланиши тил лексик ресурсларидан ижодий фойдаланиш кўламини беради.

Полисемантик сўзнинг ҳар бир лексик маъноси бир ҳалка деб олинса, улар ўзаро занжирсимон бирлашган бўлади. Уларнинг орасидан бирининг тушиб қолиши занжирнинг, яъни полисемантик сўз лексик маъноларининг ўша ердан узилиши бўлади. Омонимлар полисемантик сўз ҳисобига шу жараён асосида таркиб топишини олим М.Миртожиев алоҳида таъкидлайди.

Айрим тилшуносларнинг полисемия ҳодисасини умуман тан олмасликлари, “сўз маъносидаги кичик ўзгариш ҳам уни бошқа сўзга айлантириши” каби қарашлари ўзини оқламайди. Аксинча, кўп маъноли сўзлар синоним, антоним ва омоним каби тил тараққиётини, унинг бойлигини, кўп маънони бир сўзда сингдира олиш қобилиятини, тилдаги тежамкорлик тамойилларини яққол намоён этади. Яна бир жиҳатни алоҳида таъкидлаш керакки, кўп маъноли сўзнинг мақсадга мувофиқ маъносини танлаб ишлатиш, ўз ўрнида фойдаланиш тил эгасидан закийликни талаб этади. Ўзбек тилининг миллий корпуси учун кўп маъноли сўзларни фарқлаш дастурининг яратилиши сўз маъноларини ажратиш учун муҳим амалий аҳамият касб этади.

Полифункционаллик конверсия ҳодисаси билан боғлиқ. О.С.Ахманова “Словарь лингвистических терминов”да “конверсия – маълум бир сўзнинг ҳеч қандай аффикс олмаган ҳолда бошқа сўз туркуми вазафасида келиш ҳодисаси”³³, деб таъриф беради. Тилшуносликда полифункционаллик тушунчаси ғоят хилма-хил. Полифункционаллик тил сатҳлари бирликларининг хусусияти сифатида ажралиб туради: тадқиқотчилар морфема, сўз (сўз шакли), иборалар полифункционаллиги ҳақида фикр юритишади. Баъзан полифункционаллик торроқ, баъзан жуда кенг тушунилади.

Полифункционаллик битта грамматик синф ичида ҳам бўлиши мумкин. Масалан, ёрдамчи сўзлар орасида полифункционаллик кенг тарқалган. Айниқса, у боғловчи ва юкламалар учун характерли. Бундай сўзларнинг семантик бирлиги М.И.Черемисина томонидан исботланган: “Маълум бир ҳолатда боғловчи юклама ва аксинча, муайян позицияларда юклама боғловчи вазифасини олади”³⁴.

³³ Ахманова О.С. Словарь лингвистических терминов. – Москва, 1966. – С. 197. – 607 с.

³⁴ Черемисина М.И. Сравнительные конструкции русского языка. – Москва, 1976. – 270 с.

Тилшунослигимизда полифункционаллик масаласига доир қатор ишлар амалга оширилган. Бу муаммо таҳлиliga бағишланган Ё.Тожиев, Ж.Элтазаров, Н.Қодирова, Г.Сулаймонова, Т.Валиев, А.Ботирова ва бошқаларнинг тадқиқотлари алоҳида қимматга эга³⁵. Тил бирликларининг полифункционаллиги масаласида турли фикр-мулоҳазалар ифодаланган. Уларни умумлаштириб, қуйидаги гуруҳларга бўлиш мумкин: 1) тил бирликларига хос полифункционалликни полисемия доирасида ўрганиш; 2) полисемия ва омонимия оралиғидаги боғловчи бўғин; 3) сўз семантик структурасидаги турли сўз туркумларига хос маъноларнинг синкретизми; 4) полисемантиклик хусусияти сусайган, бироқ омонимлик даражасига етмаган бирлик.

Бобнинг “Полифункционал ва кўп маъноли сўзларни корпусда бартараф этиш учун филтрлаш тажрибасидан” деб номланган бўлимида полифункционал сўзларнинг компьютер лингвистикасидаги тадқиқи, полисемантик сўзларни филтрлаш масаласи ва тил корпусида берилиши, кўп маънолиликни бартараф этиш тажрибаси хусусида фикр юритилган. Изланиш давомида полифункционалликнинг компьютер тилшунослигида ўрганилиши, сўз маъноларини фарқлашнинг автоматик амалга оширилиши борасида бирор тадқиқот амалга оширилганлигини кузатмадик. Бу, бизнингча, назарий тилшуносликда ҳодиса борасида яқдил илмий хулосанинг мавжуд эмаслиги билан изоҳланади. Айрим кичик тадқиқотларда полисемиянинг корпусда берилиши масаласини ёритиш жараёнида бу ҳодисага йўл-йўлакай муносабат билдирилган. Жумладан, И.М.Кобозеванинг “Полисемия дискурсивных слов и возможности ее разрешения в контексте предложения (на примере слова *вот*)” мақоласида *вот* сўзи ҳам юклама, ҳам олмош вазифасида полифункционал сўз бўлиб келиши ва унинг рус миллий корпусида ифодаланиши хусусида айрим мулоҳазалар келтирилган³⁶. Полифункционал билан бир қаторда “кўп функциялилик” термини айрим тилшунослар ишида учрайди. Жумладан, А.Лефевре, “бир нечта грамматик функцияни бажарадиган лексик элемент кўп функцияли”, – дейди. Полифункционал ва кўп функцияли термини синоним сифатида қўлланади³⁷.

Ўзбек тилшунослигининг анъанавий йўналишида полифункционаллик бўйича амалга оширилган тадқиқотларга асосланиб, бундай сўзларни филтрлашда локализациянинг аҳамияти хусусида, матнда полифункционал сўзларни семантик филтрлашда лингвистик омиллар масалаларини ёритиб, семантик анализатор учун полифункционал сўзлар масаласини лингвистик жиҳатдан ечишга ҳаракат қилинди.

Корпус лингвистикасида полисемантик сўз муаммосини ечиш, полисемантик бирликни теглаш, матнни автоматик ўқиш жараёнида

³⁵Қодирова Н.А. Ўзбек тилидаги аффиксларда полифункционаллик: Филол. фан. номз. дисс. автореф. – Тошкент, 2002. – 26 б. – Б.16; Сулаймонова Г.А. Ўзбек тилида от туркумидаги сўзларнинг полифункционаллиги: Филол. фан. номз. дисс. автореф. – Тошкент, 2010. – 26 б. – Б. 8.; Элтазаров Ж. Д. Ўзбек тилида сўз туркумларининг ўзаро алоқаси ва кўчиши: Филол. фан. докт. дисс. автореф. – Тошкент, 2008. – Б. 26. – 42 б.

³⁶ Кобозева И.М. Полисемия дискурсивных слов и возможности ее разрешения в контексте предложения (на примере слова *ВОТ*) // Труды международной конференции “Диалог 2007”. – С. 250-255.

³⁷ Alan Reed Libert Hypersynonymy for Polyfunctionality // https://www.researchgate.net/publication/329881879_Hypersynonymy_for_Polyfunctionality

полисемияни бартараф этиш масаласи (“снятия полисемии”)га оид қатор тадқиқотлар вужудга келган³⁸. Лексик кўп маънолилиқ муаммосини ҳал қилишда қуйидагилар зарур: 1. Матн билан боғлиқ ҳар бир сўзнинг маъносини аниқлаш. 2. Контекстда учраган сўзга асосланиб, энг мос маъносини танлаш.

Корпусдаги кўп маъноли сўзнинг маъноси оддий изоҳли луғатдаги каби рақам билан эмас, семантик белги билан фарқланади, масалан: *арраламоқ* (тўнкани) – “жисмоний ҳаракат”, *арраламоқ* (эрни) – “нутқ”. Дастурлар семантик филтрлардан фойдаланади. Агар луғатда ҳар бир белги тегишли маънога қиёсан қўйилган бўлса, корпус матнларида сўзнинг барчасига қўйилади, шунинг учун белгини автоматик жойлаштиришда сўзнинг маъносини ажратиб бўлмайди. Кўп маънолилиқ дастури семантик филтрлардан фойдаланади. Кўп маъноли сўз гапда аниқ бир маънода ишлатилади (сўз ўйинини ҳисобга олмаганда), бу эса сўз маъносининг контекстга мос келишини англатади.

Бобнинг “Полифункционал ва кўп маъноли сўзларни ўз ичида фарқлаш омиллари” деб номланган учинчи фаслида полифункционал ва кўп маънолилиқни фарқлашда локализация ҳодисаси, полифункционал сўзларни фарқлашнинг лингвистик омиллари, кўп маъноли сўзларни фарқлаш омиллари каби масалалар ёритилган. Ўзбек тилшунослигида лексик сатҳда локаллиқнинг ифодаланишига бағишланган тадқиқот мавжуд³⁹ бўлиб, бунда асосий эътибор лексемалар семемасидаги локал семалар ва бу семаларнинг тил лексик системасидаги система ҳосил қилиш (интегротив) вазифасини очишга қаратилган. Ўзбек тилида локум ва локализация муносабати М.В.Эргашова диссертациясида ёритилган⁴⁰. Н.Ширинова диссертациясида сўз бирикмаларида предметлик ва белги-хусусият маъноларини фарқлашда релевант аҳамиятли синтактик усуллар сифатида локализация ҳодисасини кўрсатади. Предметликни маълум даражада мўътадиллаштириб, белги-хусусиятни кучайтиришнинг барча тиллар учун кенг қўлланиладиган омили тилшунослиқда локализация (жойлаштириш ёки ёнма-ён қўйиш (рус. соположение) бўлиб, бунда, одатда тобе компонент предметлик маъносини деярли йўқотиб, белги-хусусият маъносини воқелантиради. Бу усул тилимизда жуда кенг қўлланилади: *тош кўприк, олтин соат...*⁴¹

³⁸Шеманаева О.Ю., Кустова Г.И., Ляшевская О.Н., Рахилина Е.В. Семантические фильтры для разрешения многозначности в Национальном корпусе русского языка: прилагательные // Иомдин Л.Л., Лауфер Н.И., Нариньяни А.С., Селегей В.П. (ред.). Компьютерная лингвистика и интеллектуальные технологии: Труды между народной конференции «Диалог 2007». – 2007. – С. 582-587.; Толдова С.Ю., Кустова Г.И., Ляшевская О.Н. Семантические фильтры для разрешения многозначности в Национальном корпусе русского языка: глаголы // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4-8 июня 2008 г.). – Вып. 7 (14). – М.: РГГУ, 2008.; Рахилина Е.В., Ляшевская О.Н., Кобрицов Б.П., Кустова Г.И., Шеманаева О.Ю. Многозначность как прикладная проблема: Лексико-семантическая разметка в Национальном корпусе русского языка // Лауфер Н.И., Нариньяни А.С., Селегей В.П. (ред.). Компьютерная лингвистика и интеллектуальные технологии: Труды между народной конференции “Диалог 2006”, – 2006. – С. 445-450.

³⁹ Курбонова Б. Локаллик ва унинг ўзбек тилида ифодаланиши: Филол. фанлари номзоди дисс. – Тошкент, 2007. – 107 б.

⁴⁰Эргашова М. Ўзбек тилида локум ва локализация муносабати. Филол. фанлари номзоди дисс. ав-ти. – Тошкент, 2011. – 21 б.

⁴¹ Ширинова Н. Ўзбек тилида белги-хусусият ва предметлик маъноларининг фарқланишида градуонимик қаторлар. Филол. фанлари номзоди дисс. – Тошкент, 2010. – 160 б. – Б. 85.

Синтактик боғланиш, биринчидан, синтактик бирлик элементларининг ўзаро муносабатларидан иборат, яъни сўзлар ўртасидаги синтактик муносабатларни ифодалаш учун хизмат қилади. Иккинчидан, гап ва сўз бирикмасининг синтактик структурасини яратади. Учинчидан, сўзнинг лексик маъносини амалга ошириш учун шароит яратади. Сўз шакллари, ёрдамчи сўзлар, лексик элементлар, интонация, сўз тартиби синтактик алоқани юзага чиқарувчи воситалар ҳисобланади. Сўз бирикмаларида локализация (жойлаштириш) синтактик омилини оддий синтактик битишувдан фарқлаш лозимлиги С.Назарова диссертациясида таҳлил қилинган⁴². Локализация деган тушунча сўз шакллариининг жойлашувига мувофиқ келиб, полисемия ва полифункционал сўз маъноларини аниқлаш учун муҳим. Чунки муайян маънодаги сўз маълум синтактик позицияни тақозо этгани каби муайян синтактик позиция ҳам шу позицияга мос келадиган маънодаги сўзни тақозо этади. Бошқача қилиб айтганда, синтаксис қанчалик семантик табиатга эга бўлса, семантика ҳам, ўз навбатида, шунчалик синтактик табиатга эгадир⁴³.

Тил бунёдкори шахс экан, тил воситаларининг воқеланиши учун лисоний ва прагматик омиллар ҳамкорлигини назардан қочирмаслик лозим. Тадқиқот объекти бўлган полифункционал сўзлар таҳлилида ҳам лингвистик омиллар билан бир қаторда нутқий вазият, сўз қуршовининг аҳамияти инобатга олинishi керак.

ЎТОИЛда омоним деб берилган *бер, боқ, кел, ол, юбор, қол* каби феълларга полифункционал сўз сифатида қараш ўринли. Бунинг назарий асослари мазкур бобнинг биринчи фаслида таҳлил қилинди. Мустақил ва кўмакчи феъл орасидаги полифункционалликни фарқлаш омилларига тўхталсак:

Бер – феъл. Қўлига тутқизмоқ, ихтиёрига топширмоқ, таъмин қилмоқ.

Бер – кўмакчи феъл. Мустақил феъл англатган лексик маънога: 1) “бирова учун бажарганлик” грамматик маъносини қўшади. 2) “давомийлик, такрорийлик” грамматик маъносини қўшади⁴⁴.

<i>Бер</i> мустақил феъл (СБда ҳоким қисм)	
китобни, меҳрни, бахтни	бер
менга, унга, бизга	
уйда, кўчада, маҳаллада, ишда	
пулдан, ширинликдан, чойдан	
<i>Бер</i> мустақил феъл (СБда тобе қисм)	
бериб беришга берса берганди (-гунча, -гач) бергандан кейин	кел
бериб	юбор

⁴² Назарова С. Бирикмаларда сўзларнинг эркин боғланиш омиллари. Филол. фанлари номзоди дисс. – Тошкент, 1997. – 20 с.

⁴³ Гак В. Г. К проблеме семантической синтагматики. «Проблемы структурной лингвистики». – М., 1971. – С. 367.

⁴⁴ Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. – Тошкент: Ўқитувчи, 1984. – Б. 29. – 214 б.

берса берганда (-гунча, -гач) бергандан кейин	
бериб беришни (-га, -да) берса берганда (-гунча, -гач) бергандан кейин	айт
<i>Бер мустақил феъл (КФСҚнинг етакчи қисми)</i>	
бериб (-а)	кўр
	юбор
	сол

Олд бирикувчилар орқали *бер* феъли мустақил сўз сифатида воқеланган. Мустақил феъл бўлиб, келишик шакллари бошқариб, ҳоким сўз вазифасини бажарган. *Бер кел, бериб юбор* мисолларида *бериб* мустақил феъл бўлиб, битишувли алоқанинг тобе компоненти бўлиб келган.

<i>Бер кўмакчи феъл</i>	
жеркиб	бер
айтиб	
куйлаб	
текшириб	
айта	
кўра	
топшира	

Бу ўринда кўмакчи феълли сўз қўшилмасининг кўмакчи қисми вазифасида келган. Етакчи ва кўмакчи қисм равишдош шакллари орқали боғланган. Демак, мазкур феъл мустақил феъл бўлиб келганда ҳам ҳоким, ҳам тобе мавқеида воқеланади. Кўмакчи феъл бўлганда эса фақат етакчи феъл билан боғланиб, биров учун бажарилган иш-ҳаракат маъносини англатади.

Таҳлиллардан кузатилдики, полисемантик сўзларнинг маъноси морфологик ва синтактик омиллар ҳамкорлигида намоён бўлади. ЎТИЛдан полисемантик сўзларни ажратиб, лингвистик омиллар асосида бирикувчилар танлаб, мукамал база яратиш келажакда семантик анализатор учун муҳим манба бўлиб хизмат қилади.

Диссертациянинг IV боби “**Ўзбек тили семантик анализатори учун омоним, полифункционал ва кўп маъноли сўзлар қуршовини лингвистик моделлаштириш масалалари**” деб номланган. Бобнинг “Омоним сўзлар қуршовини лингвистик моделлаштириш” деб аталувчи бўлимида бир туркум, от//феъл ҳамда сифат//феъл туркуми ҳамда турли туркум доирасида вужудга келадиган омоним сўзлар қуршовини лингвистик моделлаштириш муаммолари таҳлил этилади. Компьютер тилшунослигида лингвистик моделлаштириш муҳим аҳамиятга эга. Яратилган модел асосида дастурий таъминот яратилиб, тил бирлигининг корпусда берилиши масаласи ечим топади. Ўзбек тили семантик анализатори учун омоним сўзлар

қуршовини лингвистик моделлаштириш масаласи муҳим ҳисобланади. Ҳар бир омоним турининг туркум табиатидан келиб чиқиб, уларнинг лингвистик таъминотини ва у орқали лингвистик моделини тузиш лозим. Чунки бир сўз туркуми доирасидаги омоним сўзлар қуршови от ва феъл ҳамда сифат ва феъл туркуми доирасида учрайдиган омоним сўзлар ёки турли туркум доирасида вужудга келадиган омоним сўзлар қуршовини моделлаштириш каби амалга оширилмайди. Чунки бир туркум доирасидаги омоним сўзларни грамматик шакллар орқали аниқлаб бўлмайди. Маълумки, бир туркум доирасида учрайдиган омоним сўзлар грамматик шакллар қўшилганда ҳам омонимлигини сақлаб қолади. Масалан:

Асорат I от. Тутқунлик, қуллик, эркисизлик ҳолати.

Асорат II от. Бўлиб ўтганнинг қолдиғи, излари⁴⁵.

Асорат + нинг/ни/га/да/дан = асоратнинг/ни/га/да/дан

Асорат + нинг/ни/га/да/дан = асоратнинг/ни/га/да/дан

Кўриняптики, бундай филтер орқали бир туркум доирасида учрайдиган омонимларни фарқлаш имконсиз. Шунинг учун бундай сўзлар алоҳида филтер базасини яратишни тақозо этди.

Лингвистик база бўлиб хизмат қилган “Омонимларнинг изоҳли луғати” ҳамда “Ўзбек тилининг изоҳли луғати”да бир сўз туркуми доирасидаги омонимлар сони ва турини қуйидагича аниқладик:

1) фақат от туркуми доирасидаги омоним сўзлар – 1060 та;

2) фақат сифат туркуми доирасидаги омоним сўзлар – 82 та;

3) фақат феъл туркуми доирасидаги омоним сўзлар – 235 та.

Ўзбек тилида қуйидаги усуллар орқали омонимияни бартараф этиш мумкин:

1) олдинги мустақил сўз билан бирикиш – W_P ;

2) кейинги мустақил сўз билан бирикиш – W_F .

Биз омоним сўзлар қуршовини лингвистик моделлаштиришда хусусий маънолар бирикувчанлигига асосландик. Чунки сўз серқирра ҳодиса бўлиб, бир сўз жуда кўп бирикувчилар билан қўллана олади. Сўзнинг нутқий қўлланиши ҳам ўзига анча бирикувчиларни “торта олади”. Бу ҳолда сўз семантикаси билан ишлаш, айниқса, уни моделлаштириш корпус тилшунослиги учун мураккаблигича қолаверади. Ана шу қамровни хусусий маъно (ХМ) орқали чегаралаш мақсадга мувофиқ. Бу эса семантик филтер базасини яратиш ҳамда омонимларни моделлаштиришни енгиллаштиради.

Бир туркум ичида учрайдиган омоним сўзларни семантик филтрлаш қуйидаги йўллар орқали амалга оширилиши мумкин:

1. Фақат от сўз туркуми доирасидаги омонимларни келишик шаклларини (6 та) олган ҳолда ундан сўнг қандай сўз келишига қараб (сўнг бирикувчанлик асосида) ёки от-омонимдан олдин келган сифат орқали хусусий маънолар бирикувчанлиги асосида аниқлаш.

Буни қуйидагича моделлаштириш мумкин:

⁴⁵ Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. Тошкент: Ўқитувчи, 1984. – 108 б. – Б 22.

$H_N + CS_{aff} = V$. Бунда: $H_N = \{\text{от-омоним}\}$; $CS_{aff} = \{-, \text{-нинг}, \text{-ни}, \text{-га}, \text{-да}, \text{-дан}\}$; $V = \{\text{феъл сўз}\}$.

$H_N + CS_{aff} = N/Num/Adj/Adv/Pr/Mim$. Бунда: $H_N = \{\text{от-омоним}\}$; $CS_{aff} = \{-, \text{-нинг}, \text{-ни}, \text{-га}, \text{-да}, \text{-дан}\}$; $N = \{\text{от}\}$, $Num = \{\text{от}\}$, $Adj = \{\text{сифат}\}$, $Adv = \{\text{равиш}\}$, $Pr = \{\text{олмош}\}$, $Mim = \{\text{таклид}\}$.

$Adj + H_N$. Бунда: $Adj = \{\text{сифат}\}$, $H_N = \{\text{от-омоним}\}$. Мана бу модел асосида ишлайдиган от-омонимлар ҳам лексиконда мавжуд бўлиб, булар ўзидан олдин келган сифат билан бирикиб, бирикувчиси отнинг сифатловчиси бўлиб, от-омоним маъносини аниқлаб келади.

Булар фақат от сўз туркуми доирасидаги омонимларни аниқлаш учун ишлаб чиқилган инвариант модел бўлиб, бу бир қатор субмоделларда воқеланади.

Фақат сифат сўз туркуми доирасида учрайдиган омонимларни ундан кейин кела оладиган хусусий маънолар бирикувчанлиги асосида сўнг усул орқали аниқлаш.

$H_{adj} + N$. Бунда: $H_{adj} = \{\text{сифат-омоним}\}$; $N = \{\text{от сўз}\}$;

$H_{adj} + V$. Бунда: $H_{adj} = \{\text{сифат-омоним}\}$; $V = \{\text{феъл сўз}\}$.

Фақат феъл туркумидаги омонимларни фарқлаш қуйидаги йўллар билан амалга ошади:

1) феъл-омоним 2-шахс буйруқ майлида ёки феълнинг ноаниқ формаси (-моқ)да бўлса, ўзидан олдин хусусий маънолар бирикувчанлиги асосида сўнг усул билан аниқланади.

2) агар феъл-омоним феълнинг вазифа шакллари қўшимчалари (равишдош, сифатдош, ҳаракат номининг -ш, -иш, -в, -ув шакли) билан келган бўлса, феъл-омоним ё ўзидан олдин, ё кейин келадиган сўз билан бирикма ҳосил қилади.

Буларни қуйидагича моделлаштириш мумкин:

$WG + H_V$. Бунда: $WG = \{\text{от}, \text{равиш}, \text{равиш характеридаги сифат}, \text{сон}, \text{олмош}\}$; $H_V = \{\text{феъл-омоним}\}$

$Y_r^f + O_f^n$. Бунда: $Y_r^f = \{\text{етакчи феъл: -б/-б, -а/-й}\}$; $O_f^n = \{\text{феъл-омоним}\}$.

Булар фақат феъл сўз туркуми доирасидаги омонимларни аниқлаш учун ишлаб чиқилган инвариант модел бўлиб, бу бир қатор субмоделларда воқеланади.

От ва феъл ҳамда сифат ва феъл туркуми доирасида учрайдиган 120 та омоним сўзларни лингвистик базадан ажратиб олдик. Бу омоним сўзларнинг 102 таси от ва феъл, 18 таси сифат ва феъл сўз туркуми орасида воқеланган.

Ари I от. Парда қанотли чақадиган ҳашарот.

Ари II феъл. Йўқолмоқ, тугамоқ, кетмоқ⁴⁶.

1) ари+ни/лар/ча=от

2) ари+ма/ган/гунча/гин/май/ди=феъл

Ая I от. Фарзандни дунёга келтирган аёл, она.

Ая II феъл. Авайламоқ, эҳтиёт қилмоқ⁴⁷.

⁴⁶ Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. Тошкент: Ўқитувчи, 1984. – 108 б. – Б 21.

⁴⁷ Ўша манба. – Б. 24.

1) ая+жон/м/нг/си/миз/нгиз/нинг/ни/га/да/дан/лар=от

2) ая+ ма/май/мас/ган/р/б/гач/гунча/гудек/й/нг/син/нгиз =феъл.

От ва феъл орасидаги юқоридаги каби омонимларни мана шундай филтер орқали фарқладик.

От-омонимни отга хос луғавий ва синтактик шакл ясовчи қўшимчалар ёрдамида аниқлаш мумкин. Буни қуйидагича моделлаштириш мумкин:

$$H_N + LF_N + SF_N = N$$

Бунда H_N – омоним сўз, LF_N (Lexical forms of the noun) – отнинг луғавий шакллари, SF_N (Syntactic forms of the noun) – отнинг синтактик шакллари, N (Noun) – от сўз туркуми.

Феъл-омонимни эса феълга хос луғавий ва синтактик шакл ясовчи аффикслар ёрдамида аниқлаш мумкин. Буни эса қуйидагича моделлаштирамиз:

$$H_V + LF_V + SF_V = V$$

Бунда H_V – омоним сўз, LF_V (Lexical forms of the verb) – феълнинг луғавий шакллари, SF_V (Syntactic forms of the verb) – феълнинг синтактик шакллари, V (Verb) – феъл сўз туркуми. Бу моделлар бир қатор субмоделларда вокеланади.

Тадқиқотда лингвистик база бўлиб хизмат қилган “Ўзбек тили омонимларининг изоҳли луғати”да турли туркум доирасида вужудга келадиган омоним сўзлар ранг-баранг бўлиб, уларнинг қуршовини лингвистик моделлаштириш индивидуал филтер ишлаб чиқишни тақозо этди. Чунки турли туркум доирасига икки, уч, тўрт туркум ҳам тегишли бўлиб, айрим омонимлар икки ва ҳатто беш-олтитагача маънога эгалигини кузатиш мумкин. Омонимларнинг ранг-баранг туркум таснифидан кўринадики, турли туркум доирасидаги омоним сўзларни фарқлаш учун ҳар бир сўз маъносига мувофиқ филтер базаси талаб қилинади.

Бобнинг иккинчи бўлими “Полифункционал сўзлар қуршовини лингвистик моделлаштириш” деб аталади. Полифункционал сўз масаласи жаҳон тилшунослигида тўлиқ ечим топмаган бўлиб, унинг автоматик аниқланиши, корпусда бартараф этилиши, лингвистик филтри ва модели борасида бирор тадқиқот учратмадик. Мазкур фаслда ўзбек тилида мавжуд назарий қарашларни умумлаштириб, улардан фойдаланиб, семантик анализатор учун полифункционал сўзларни моделлаштириш борасидаги дастлабки ишларни бажаришга ҳаракат қилинди.

Олтин, кумуш, чўян, касал, темир, ёғоч, чўлоқ, қуйи, бўртма каби бир қатор от-сифат орасидаги полифункционал сўзлар қуйидагича моделлаштирилади:

$P_i^N + V$. Бунда: $P_i^N = \{\text{От бўлиб келган полифункционал сўз, } i=1..n\}$;
 $V = \{\text{феъл}\}$.

$P_i^N + V = \{\text{олтин топмок/ковламок/сотиб олмок/бермок...}\}$

$P_1^{ADJ} + N$. Бунда: $P_{ADJ} = \{\text{Сифат бўлиб келган полифункционал сўз, } i=1..n\}$; $N = \{\text{от}\}$

$P_1^{ADJ} + N = \{\text{олтин узук/билакузук/соат/бош/кўл...}\}$

Бобнинг учинчи фасли “Кўп маъноли сўзлар қуршовини лингвистик моделлаштириш” деб номланади. Маълумки, лексик кўп маънолиликни (инглизча *word sense disambiguation, WSD*) бартараф қилиш – бу полисемик сўз ёки иборанинг маъносини танлаш вазифасидан иборат бўлган табиий тилни қайта ишлашнинг ҳал қилинмаган муаммосидир. Бу муаммо дискурсив таҳлилда, натижаларнинг кидирув тизимлари томонидан долзарблигини оптималлаштиришда, анафорик мурожаатларни ҳал қилишда, матннинг лингвистик боғлиқлигини ўрганишда, хулосаларни таҳлил қилишда юзага келади. Кўп маънолиликни фарқлашда ҳам бирикувчилик тамойилига таянилади. Сўзнинг қайси маънода келгани ундан олдин ёки кейин келган сўз орқали аниқланади.

Ўзбек тилининг ички имкониятлари ҳамда хусусиятларини инобатга олиб, кўп маъноли сўзларни моделлаштиришнинг қуйидаги йўллари кўрсатиш мумкин (ёқа сўзи мисолида):

$$N_j^{ps.1} + PS_1$$

$N_j^{ps.1}$ = полисемантик сўзнинг дастлабки (ўз) маъноси билан бирикувчи от туркумидаги сўзлар.

$$N_j^{ps.1} = \{\text{кўйлак, кийим, ...}\}; j = 1 \dots n$$

PS_1 = дастлабки (ўз, денотатив) маъносида келган полисемантик сўз.

Масалан: *кўйлак ёқаси, кийим ёқаси*

$$PS_1 + V_j^{ps.}$$

$V_j^{ps.}$ = полисемантик сўзнинг дастлабки (ўз) маъноси билан бирикувчи феъл туркумидаги сўзлар.

$$V_j^{ps.} = \{\text{тикмоқ, очмоқ, узмоқ, ...}\}; j = 1 \dots n$$

Масалан: *ёқа очмоқ, ёқа тикмоқ*

$$Adj_j^{ps.} + PS_1$$

$Adj_j^{ps.}$ = полисемантик сўзнинг дастлабки (ўз) маъноси билан бирикувчи сифат туркумидаги сўзлар.

$$Adj_j^{ps.} = \{\text{кенг, тор, узун, калта ...}\}; j = 1 \dots n$$

Масалан: *кенг ёқа, тор ёқа*

Полисемантик сўз бошқа (кўчма, коннотатив) маъноларда келганда эса қуйидагича моделлаштирилади:

$$N_j^{ps.2} + PS_2$$

$N_j^{ps.2} + PS_2$ = полисемантик сўзнинг кўчма маънолари билан бирикувчи от туркумидаги сўзлар.

$$N_j^{ps.2} = \{\text{арик, кўл, ўлим, канал, кўча, йўл, ...}\}; j = 1 \dots n$$

PS_2 = полисемантик сўзнинг кўчма маънода қўлланишини ифодалайди.

Масалан: *арик ёқаси, кўл ёқаси, каналнинг ёқаси, кўча ёқаси, йўлнинг ёқаси, тепалик ёқаси*

Моделлаштириш табиий тилни қайта ишлашнинг асосий қуроли⁴⁸. Семантик анализатор учун омоним, полисемантик ва полифункционал сўзларни моделлаштиришда бир қатор муаммолар мавжуд. Биринчидан, сўз

⁴⁸Хамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. Монография. – GlobeEdit, 2020. – 253 б. – Б. 197.

семантикаси билан ишлаш анча мураккаб жараён. Чунки сўз серқирра ҳодиса. Иккинчидан, ўзбек тилидаги изоҳли луғатда айрим сўзлар маъносининг берилишида, изоҳида чалкашликлар мавжудки, бу масалани янада чигаллаштиради. Семантик анализатор учун мукамал лингвистик база яратиш учун ҳар бир сўз маъноси устида юқорида тузилган моделлардан фойдаланган ҳолда индивидуал ишлаш талаб этилади. Бунинг учун сўз маънолари, изоҳи, туркуми, қандай характер касб этиши мукамал ҳолатда маълумотлар омборига киритилиши лозим. Семантик анализатор ишда таклиф этилаётган қолиплар асосида сўз маъносини аниқлашга, мисоллар массивини кўрсатишга хизмат қилади.

ХУЛОСА

1. Лингвистик анализатор – автоматлашган қидирув тизимида табиий тилни қайта ишлаш жараёнида сўзнинг морфологик, синтактик ва семантик хусусиятини ифодалайдиган восита. Бу воситалар ўз ичига лемматизатор, морфоанализатор, стеммер, парсер, орфоқорректорларни олади. Бу анализаторларнинг ҳар бири ишлаш принципи, мақсади, вазифасига эга.

Семантик анализаторни ишлаб чиқиш учун даставвал семантик анализ ва синтезнинг матннинг автоматик таҳлилидаги аҳамиятини ёритиш муҳим. Автоматик семантик таҳлил компьютер лингвистикасининг актуал ва шу билан бирга, мураккаб вазифаларидан биридир. У инсон ақл-заковати билан боғлиқ бўлган ҳар қандай тизимни яратишнинг зарурий босқичидир.

2. Дунё тилшунослигида морфологик ва синтактик анализатор бўйича кенг тадқиқотлар олиб борилган, аммо семантик анализаторни яратиш бўйича махсус амалий иш амалга оширилмаган. Чунки семантикани расмийлаштириш ўта мураккаб вазифа. Муаммонинг туб мураккаблиги шундаки, мазмун (маъно)ни ўрганишда тилдан ташқарига чиқиб, ташқи оламга, тил доирасидан ташқарида ётувчи предмет ва тасаввурлар таснифига муурожаат қилишга тўғри келади.

3. Автоматик семантик анализ ва синтез компьютер лингвистикасининг мураккаб вазифаларидан бири бўлиб, матннинг автоматик таҳлилидаги аҳамияти катта. Ахборот-қидирув тизимида семантик анализ ва синтез жараёнини мукамаллаштириш, семантик қидирув ва унинг муаммоларини ечиш семантик анализатор вазифасини аниқлаштиришга, келажакда СемА (семантик анализатор) тузишга йўл очади. СемА ишлаши учун дастлаб АҚТда семантик анализ ва синтез жараёни тизимли йўлга қўйилиши талаб этилади. Ўзбек тили СемАнинг лингвистик таъминотини яратиш, семантик анализатор ва унинг ахборот-қидирув тизимида тутган ўрнини аниқлаш бевосита семантик анализ ва синтезнинг матннинг автоматик таҳлилидаги аҳамиятига боғлиқ.

4. Омонимияни бартараф этиш масаласини ҳал қилиш омоним сўзларнинг компьютер лингвистикасидаги тадқиқи ва уларни семантик таҳлил қилиш тажрибасини, омонимларни таҳлил қилиш усулларини пухта ўрганиш, улардан ўринли амалий фойдаланиш ҳамда омонимларни туркум асосида таснифлаш лозимлигини талаб этади. Ишда омоним сўзлар бир сўз

туркуми доирасидаги омонимлар (1); от ва феъл ҳамда сифат ва феъл туркуми орасида кузатиладиган омонимлар (2); турли туркум доирасида вужудга келадиган омонимлар (3) каби гуруҳларга бўлинди. Айнан шундай тасниф қилинишининг лингвистик асоси мавжуд. Чунки омонимларнинг филътрини яратиш учун уларни фарқловчи омилларнинг ўрни ва аҳамияти муҳим.

5. Тил бирликлари лексик, морфологик, синтактик омиллар билан биридан фарқланади. Бирликларни фарқлашда ҳар бир омил ўз вазифасига эга. Қайсидир омил етакчилик қилса, бошқа омил кучсизланиши кузатилиши мумкин ёки омиллар ҳамкорлиги (бир неча омил биргаликда таъсир этиши) асосида фарқланиши кузатилади. Тадқиқот натижасида омоним сўзни фарқлашда морфологик ва синтактик омил муҳим ўрин тутиши аниқланди. От ва феъл орасидаги омонимликда морфологик; сифат ва феъл орасидаги омонимликда морфологик-синтактик; фақат от ва фақат феъл туркуми доирасидаги омонимликда синтактик-морфологик; фақат сифат туркуми доирасидаги омонимликда синтактик; турли туркум доирасидаги омонимликда эса морфологик, синтактик, морфологик-синтактик ва синтактик-морфологик омил етакчилик қилиши аниқланди.

6. Ўзбек тили семантик анализатори учун омоним сўзлар маълумотлар базасини шакллантириш учун тилда мавжуд бўлган омоним сўзларни “Омонимларнинг изоҳли луғати” ҳамда “Ўзбек тили изоҳли луғати” (5 жилдди)дан йиғиб, туркум асосида таснифлаб, алифбо тартибида изоҳи билан бериш (1); морфологик йўл билан фарқланадиган омонимлар учун қўшимчалар комбинациясини ишлаб чиқиш (2); ўнг ва сўнг бирикувчилар билан фарқланадиган омонимлар учун синтактик филътр базасини яратиш (3); яратилган филътр базаси асосида омонимларнинг лингвистик моделларини ишлаб чиқиш(4); лингвистик моделлар орқали математик моделлар яратиш (5) каби тавсиялар ишлаб чиқилди.

7. Полифункционал ва полисемантик сўзларнинг маъносини фарқлашда локализация усули иш беради. Бу ҳодисаларни фарқлашда морфологик ва синтактик омиллар муҳим ўрин тутди. Бирикувчанлик тамойили асосида морфологик шакллар ўрни муҳимлиги ҳам намоён бўлди.

Омонимларнинг маъносини фарқлашда коллокация ҳодисаси муҳим омил сифатида хизмат қилди. Коллокацион бирикмалар орқали омонимларнинг маъносини топиш ва унинг қўлланиш частотасини аниқлаш мумкин. Омоним сўз коллокация таркибида омонимлигини йўқотади. Чунки бирикма таркибида фақат битта маъноси билан реаллашади. Коллокация орқали сўзлар ва сўз бирикмалардан қанча кўп фойдаланиш ҳақида статистик маълумот олиш мумкин. Бу орқали эса корпус тилшунослиги кучли ва кучсиз бирикишларни аниқлай олади.

8. Омоним, полисемантик ва полифункционал сўзларни бартараф этиш компьютер тилшунослигининг кўплаб дастурларида, хусусан, қидирув тизимларида аҳамиятли. Чунки у муайян сўров синфларини қайта ишлаш аниқлигини ошириши ёки сақланган ахборот миқдорини камайтириши мумкин. Таҳлиллардан кўринадикки, омоним, полисемантик ва

полифункционал сўзларни бартараф этиш учун Яширин Марков модели етакчилик қилади. Бу модел омонимларни эҳтимолий ёндашув асосида аниқлашга қаратилганлиги билан аҳамиятлидир.

9. Тадқиқот давомида бир туркум доирасидаги омонимларни аниқлаш учун 7 та инвариант модел ва 6 та субмодел; от ва феъл ҳамда сифат ва феъл туркуми доирасида учрайдиган омоним сўзлар 4 та инвариант ва 2 та субмодел; турли туркум доирасида вужудга келадиган омоним сўзлар қуршови эса 20 та инвариант модел, 3 та субмодел орқали аниқланди. Ўзбек тилида баъзи полифункционал сўзларни аниқлаш учун 12 та модел ва 37 та субмодел ишлаб чиқилди. Айрим кўп маъноли сўзларни аниқлаш учун 10 модел тузилди.

10. Таҳлиллардан кўриндики, семантик анализатор учун омоним, полисемантик ва полифункционал сўзларни моделлаштиришда бир қатор муаммолар мавжуд. Чунки сўз серқирра ходиса бўлиб, маъно билан ишлаш мураккаб жараён. Ўзбек тилидаги изоҳли луғатда айрим сўзлар маъносида чалқашликлар мавжудки, бу масалани янада қийинлаштиради. Семантик анализатор учун мукаммал лингвистик база яратиш учун ҳар бир сўз маъноси устида индивидуал ишлаш талаб этилади. Бунинг учун сўз маънолари, изоҳи, туркуми, қандай характер касб этиши мукаммал ҳолатда маълумотлар омборига киритилиши лозим.

**НАУЧНЫЙ СОВЕТ DSc.03/30.12.2019.Fil.05.02 ПО ПРИСУЖДЕНИЮ
УЧЕНЫХ СТЕПЕНЕЙ ПРИ ФЕРГАНСКОМ ГОСУДАРСТВЕННОМ
УНИВЕРСИТЕТЕ**

**ТАШКЕНТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ УЗБЕКСКОГО ЯЗЫКА
И ЛИТЕРАТУРЫ**

ГУЛЯМОВА ШАХНОЗА КАХРАМОНОВНА

**ЛИНГВИСТИЧЕСКИЕ ОСНОВЫ СЕМАНТИЧЕСКОГО
АНАЛИЗАТОРА УЗБЕКСКОГО ЯЗЫКА**

10.00.11 – Теория языка, прикладная и компьютерная лингвистика

АВТОРЕФЕРАТ

докторской диссертации (DSc) по филологическим наукам

Фергана – 2022

Тема диссертации на соискание ученой степени доктора наук (DSc) зарегистрирована за номером B2021.2.DSc/FH325 в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан.

Диссертация выполнена в Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои.

Автореферат диссертации на трех языках (узбекском, русском, английском (резюме)) размещен на веб-странице Научного совета (www.fdu.uz) и на информационно-образовательном портале «ZiyoNet» (www.ziynet.uz).

Научный консультант:

Менглиев Бахтиёр Ражабович
доктор филологических наук, профессор

Официальные оппоненты:

Самситова Луиза Хамзиевна
доктор филологических наук, профессор

Улуков Носиржон Мухаммадалиевич
доктор филологических наук, профессор

Шахабитдинова Шохдила Хашимовна
доктор филологических наук, профессор

Ведущая организация:

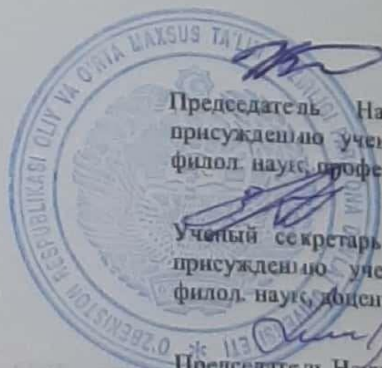
Каршинский государственный университет

Защита диссертации состоится «1» март 2022 года в «10» часов на заседании Научного совета DSc.03/30.12.2019.Fil.05.02 при Ферганском государственном университете (Адрес: 150100, г. Фергана, ул. Мураббийлар, дом 19. Тел.: (99873) 244-44-02; Факс: (99873) 244-44-93; Web-site: www.fdu.uz; e-mail: fardu_info@umail.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Ферганского государственного университета (зарегистрирована за номером 150). (Адрес: 150100, г. Фергана, ул. Мураббийлар, дом 19. Тел.: (99873) 244-44-02; Факс: (99873) 244-44-93; Web-site: www.fdu.uz; e-mail: fardu_info@umail.uz).

Автореферат диссертации разослан «8» февраль 2022 года.

(Реестр протокола рассылки № 44 от «9» февраль 2022 года).



Ш.М.Искандарова
Председатель Научного совета по
присуждению ученых степеней, доктор
филол. наук, профессор

М.Т.Зокирова
Ученый секретарь Научного совета по
присуждению ученых степеней, канд.
филол. наук, доцент

А.Б.Мамажонов
Председатель Научного семинара при
Научном совете по присуждению ученых
степеней, доктор филол. наук, профессор

ВВЕДЕНИЕ (Аннотация (DSc) докторской диссертации)

Актуальность и необходимость темы диссертации. В мировом языкознании прикладная лингвистика в 1930-1940-е годы понималась прежде всего через описание процесса обучения иностранному языку, методики его преподавания, специфики грамматики. В России прикладная лингвистика с 1950-х годов связана с вычислительной техникой и системами автоматической обработки информации. Прикладная лингвистика, сформировавшаяся в 1960-е годы как специальное научное направление, ориентирована на решение практических вопросов использования языка. В частности, корпусная лингвистика, появившаяся в мировой лингвистике в 60-е годы XX века, создала новую методологическую основу речевой практики, приоритетные направления исследований. Корпус языка и связанная с ним наука, а именно корпусная лингвистика, это новое направление, которое в конце XX – начале XXI века бурно вошло в языкознание и методику преподавания языка. Создание корпуса языка и развитие корпусной лингвистики является одним из актуальных вопросов в области лингвистики, методики преподавания языка. Современные условия языковой деятельности предполагают изучение не только речевых образцов, но и отдельных хронологических ситуаций, систематизацию речевой практики.

В мировой лингвистике в конце XIX – начале XX века корпуса стали создаваться либо для лингвистических исследований, либо с целью решения часто практических задач (например, для расчета частотности языковых единиц). Для корпусной лингвистики с изобретением и широким применением компьютеров начался новый этап развития, и анализ отличий созданного корпуса от существующего не только в формате хранения, но и в объемах стал актуальным вопросом. В области компьютерной лингвистики создание различных целевых корпусов, совершенствование, обогащение и расширение существующих корпусов – одна из проблем, которая ждет решения. Компьютерная лингвистика как область применения средств вычислительной техники может быть использована для организации программ, технологий обработки информации и моделирования языковой деятельности в конкретных условиях, а также в лингвистике в рамках компьютерных моделей языка и смежных с ними дисциплин. В связи с тем, что язык является достаточно сложным образованием, в компьютерной лингвистике развиваются различные направления, такие как автоматический анализ текста; автоматический синтез текста; создание и поддержка автоматических словарей; создание автоматизированных информационно-поисковых систем; машинный перевод; создание систем автоматического изучения языка; создание языковых баз данных; разработка программных средств решения теоретических и прикладных лингвистических задач.

За годы независимости в Узбекистане были проведены многочисленные исследования по компьютерной и корпусной лингвистике. Следовательно, что “в сегодняшнюю эпоху глобализации каждый народ, каждое независимое государство должны уделять приоритетное внимание обеспечению своих

национальных интересов, сохранению и развитию, прежде всего, своей культуры, древних ценностей, родного языка. В целях коренного повышения авторитета узбекского языка в общественной жизни нашего народа и на международном уровне, воспитания подрастающей молодежи в духе патриотизма, преданности национальным традициям и ценностям, преемственности богатому наследию наших великих предков, обеспечения полноценного внедрения государственного языка в нашей стране»⁴⁹ корпусная лингвистика как перспективное научное направление, требует решения ее теоретических проблем. Это ставит на повестку дня одну из актуальных задач, стоящих перед нашей наукой, – создание лингвистических основ разграничения омонимичности, многозначности и полифункциональности в автоматическом анализе единиц узбекского языка.

В определенной степени данное исследование послужит ориентиром при реализации задач, поставленных на указах Президента Республики Узбекистан. В особенности Указ Президента Республики Узбекистан от 13.05.2016 года №УП-4797 «Об организации Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои», Указ Президента Республики Узбекистан от 07.02.2017 года «О стратегии действий по дальнейшему развитию Республики Узбекистан», Постановление Президента Республики Узбекистан от 17.02.2017 года № ПП-2789 «О мерах по дальнейшему совершенствованию деятельности академии наук, организации, управления и финансирования научно-исследовательской деятельности», Постановление Президента Республики Узбекистан от 04.10.2019 года № ПП-4479 «О широком праздновании тридцатилетия принятия закона Республики Узбекистан «О государственном языке», Указ Президента Республики Узбекистан от 21.10.2019 года № УП-5850 «О мерах по кардинальному повышению роли и авторитета узбекского языка в качестве государственного языка», Послание Президента Республики Узбекистан Шавката Мирзиёева Олий Мажлису от 24.01.2020 года, Указ Президента Республики Узбекистан от 20.10.2020 года № УП-6084 «О мерах по дальнейшему развитию узбекского языка и совершенствованию языковой политики в стране» и других нормативных правовых актах.

Связь диссертационного исследования с приоритетными направлениями развития науки и технологий республики. Диссертационная работа выполнена в соответствии с приоритетным направлением развития науки и технологий республики «Социальное, правовое, экономическое, культурное, духовно-просветительское развитие информационного общества и демократического государства, развитие инновационной экономики».

Обзор зарубежных исследований по теме диссертации⁵⁰. Научные исследования, направленные на изучение лингвистического обеспечения

⁴⁹ Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг «Ўзбек тилининг давлат тили сифатидаги нуфузи ва мавқеини тубдан ошириш чора-тадбирлари тўғрисида»ги фармони // www.xabar.uz

⁵⁰ Диссертация мавзуси бўйича хорижий илмий тадқиқотлар шарҳи google.scholar.com., cheloveknauka.com, dslib.net, reseachgate. www.kaznu.kz, bloomsbury.com, www.utexas.edu, english.pku.edu.cn, www.lang.ox.ac.uk,

автоматизированных информационно-поисковых систем в мировой лингвистике, проводятся в ведущих исследовательских центрах и высших учебных заведениях, в том числе, Oxford University (Великобритания), Cambridge University (Великобритания), Humboldt Universitat (Германия), Boston University, University of New York, University of Pennsylvania, University of California (США); Indiana University (США), Парижский университет Сорбонна (Universite de Paris (Франция), Universitat de Valencia (Испания), University of Zurich (Швейцария), университет Упсалла (Швеция), Kiofo University (Япония), Australian National University (Австралия), Стамбульский технический университет (Турция), Московский государственный университет (Россия), Московский государственный лингвистический университет (Россия), Санкт-Петербургский государственный университет (Россия), Уральский государственный педагогический университет, Харьковский государственный педагогический университет (Украина), Bakı Dövlət universiteti (Азербайджан), Институт прикладной семиотики Академии наук Республики Татарстан (Республика Татарстан, Россия), Казанский федеральный университет (Республика Татарстан, Россия), Таврическая академия Крымско федерального университета имени В.И.Вернадского, (Крымская Республика, Россия); Национальный университет Узбекистана имени Мирзо Улугбека, Ташкентский государственный университет узбекского языка и литературы имени Алишера Навои, Ташкентский университет информационных технологий имени Мухаммада аль-Хорезми и его филиалы, Бухарский государственный университет, Термезский государственный университет (Узбекистан).

В мировой лингвистической информационно-поисковой системе (ИПС) были получены следующие научные достижения в результате исследования семантического анализатора в информационно-поисковой системе, технологий и инструментов автоматической обработки текста, программ автоматического анализа/синтеза текста, семантических анализаторов: понятие лингвистического анализатора, его виды, освещено значение семантического анализа и синтеза в автоматическом анализе текста (Boston University, New York University, АКШ; Уральский федеральный Университет, Россия); определено значение и методы использования лингвистического и синтаксического парсера (Уральский федеральный Университет, Россия); изучены технологии автоматической обработки текста (University of Tokyo, Япония, Australian National University, Австралия, Oxford University Language Centre, Англия); разработаны основы таких аналитических систем, как Морфо-Анализатор, Сегморф, Линтез, Пси-Ан (Харьковский государственный педагогический университет, Украина; Казанский федеральный университет (Россия, Республика Татарстан); модель и программы автоматической морфологической и синтаксической обработки

текста (Харьковский государственный педагогический университет, Украина; Бакинский государственный университет, Азербайджан); в компьютерной лингвистике тюркских языков исследованы вопросы компьютерной обработки естественного языка, построения синтаксико-семантического анализатора (Институт прикладной лингвистики и семиотики АН РТ, Казанский федеральный университет (Россия, Республика Татарстан), Таврическая академия Крымский федеральный университет имени В.И.Вернадского (Крымская Республика, Россия); разработаны лингвистическое обеспечение узбекско-английского машинного перевода, принципы формирования авторского корпуса узбекского языка и его лингвистического обеспечения, проблемы графического анализа единиц узбекского языка, принципы формирования лингвистической базы языкового корпуса; разработаны теоретические основы семантического тегирования единиц узбекского языка, изучены основы создания инструментальных средств узбекско-английского параллельного корпуса, лингвистические основы создания учебного корпуса узбекского языка, вопросы создания базы идиом в узбекском национальном корпусе, морфоанализатора узбекского языка и его лингвистического обеспечения (Национальный университет Узбекистана имени Мирзо Улугбека, Ташкентский государственный университет узбекского языка и литературы имени Алишера Навои, Ташкентский университет информационных технологий имени Мухаммада аль-Хорезми, Термезский государственный университет, Бухарский государственный университет).

В мировой лингвистике ведутся исследования по созданию автоматических анализаторов естественного языка, в том числе программ для анализа информационно-поисковых систем, языковых корпусов для обработки средствами современных информационных технологий, семантического анализатора и их лингвистических основ.

Степень изученности проблемы. В мировой компьютерной лингвистике ряд ученых провели научные исследования по семантическому анализу и анализатору. В качестве таких работ можно отметить исследования следующих ученых, как А.В.Тузова, М.В.Мозговой, А.В.Сокирко, Н.А.Счлаефера, А.В.Мочаловой, Д.Ш.Сулейманова, А.Р.Гатиатуллина, Б.Ергеша, А.Шарипбайя, Г.Бекмановой, С.Липницкого, Э.А.Каневского, К.К.Боярского⁵¹.

⁵¹ Тузов, В.А. Компьютерная семантика русского языка / В.А. Тузов. – СПб.: Изд-во С.-Петерб. ун-та, 2004. – 400 с.; Мозговой, М.В. Простая вопросно-ответная система на основе семантического анализатора русского языка / М.В. Мозговой // Вестник СПб университета. – 2005. – сер. 10. – вып. 1. – С. 116–122; Сокирко, А.В. Семантические словари в автоматической обработке текста (по материалам системы ДИАЛИНГ): дис. ... канд. техн. наук: 05.13.17 / Сокирко Алексей Викторович. – М., 2001. – 120 с.; Schlaefer, N. A Semantic Approach to Question Answering / N. Schlaefer. – Paperback, 2007. – 96 p.; Мочалова, А.В. Алгоритм семантического анализа текста, основанный на базовых семантических шаблонах с удалением / А.В. Мочалова // Научно-технический вестник информационных технологий, механики и оптики. – 2014. – № 5. – С. 126–132.; Mathematical model of an ontological-semantic analyzer using basic ontological-semantic patterns / A.V.Mochalova, V.A.Mochalov // Lecture Notes in Artificial Intelligence, Proceedings of 15th Mexican International Conference on Artificial Intelligence. – 2016. – P. 53–66.; Семантический анализ. Некоторые проблемы автоматической обработки текста / А.В.Мочалова, В.П.Захаров, В.А.Кузнецов, В.А.Мочалов. –

В мировой и тюркской лингвистике широко изучены вопросы составления морфологического/синтаксического анализатора. В частности, это подробно изложено в исследовании ряда таких ученых, как А.В.Дибо, А.К.Асриян, А.В.Шеймовича, А.Р.Гатиатуллина, А.М.Баширова, П.В.Кельтова, Н.А.Исраиловой, П.С.Бакасовой, М.Г.Мальковского, А.С.Старастина, М.Г.Мальковского, Д.Н.Ёунгера, Е.А.Каневского, Н.В.Колпаковой, П.С.Бакасовой, С.Берда, Э.Клейна, Э.Лопера, М.Коробова, А.Д.Москвиной, Д.Орловой, П.В.Паничевой, О.А.Митрофановой, О.Н.Ляшевской, И.Астафьевой, А.Бонч-Осмоловской, А.Гарейшиной, Ю.Гришиной, В.Дьячкова, М.Ионова, А.Королевой, М.Гудринского, А.Литягиной, Е.Лучиной, Е.Сидоровой, С.Толдовой, С.Савчука, С.Коваль и К.К.Боярского⁵².

Петрозаводск: ПИН, 2016. – 60 с.; Семантический анализатор русскоязычного текста для вопросно-ответной системы // ДИССЕРТАЦИЯ на соискание ученой степени кандидата технических наук. – Петрозаводск, 2017. – 128 с.; Сулейманов Д.Ш., Гатиатуллин А.Р. Структурно-функциональная компьютерная модель татарских морфем. — Казань: Фэн, 2003. — 220 с. – С. 152.; Yergesh B, Sharipbay A, Bekmanova G, Lipnitskii S. Sentiment analysis of kazakh phrases based on morphological rules // Известия Кыргызский государственный технический университет, теоретический и прикладной научно-технический журнал 2016 № 2(38). – С. 39-43.; Каневский Е.А., Боярский К.К. Семантико-синтаксический анализатор SemSin // Компьютерная лингвистика и интеллектуальные технологии. 2012; Боярский К.К., Каневский Е.А. Предсинтаксический модуль в анализаторе SemSin // Интернет и современное общество. Санкт-Петербург, 2013. С. 280–286.

⁵² Асриян А.К. Сравнение инструментов морфологической разметки / Intellectual potential of the XXI century 2017 / <http://www.sworld.education>; Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов тюркских языков // Филология и культура. Philology and culture. 2014. №2(36). – С. 20-26.; Гатиатуллин А.Р., Баширов А.М. Морфологический анализатор тюркских словоформ на базе структурнофункциональной модели тюркской морфемы // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». –Казань, 2017. – 327 с.; Гатиатуллин А. Р., Баширов А.М., Осипов Г.С., Смирнов И.В., Шелманов А.О. Методы лингвистического анализа текстов на татарском языке и их применение в поисковой системе Eхactus // Труды ИСА РАН. Том 66. 1/2016. – 18-25 с.; Желтов П.В. Разработка морфологического анализатора чувашского языка // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Казань, 2017. – 327 с.; Желтов П.В. Создание национального корпуса чувашского языка: проблемы и перспективы // Современные проблемы науки и образования. – 2015. – № 1-1.;

URL: <http://www.science-education.ru/ru/article/view?id=19046> (дата обращения: 18.01.2021); Исраилова Н.А., Бакасова П.С. Морфологический анализатор кыргызского языка // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Казань, 2017. – 327 с.; Мальковский М.Г., Старостин А.С. Модель синтаксиса в системе морфо-синтаксического анализа «TREEON» //Тр. междунар. Конф. Диалог 2006. – М., 2006. – С. 481-492.; Younger D.H. Recognition and parsing of context-free languages in time n3. Information and Control 10 (2). – P.189-208.; Каневский Е.А., Колпакова Н.В. К вопросу построения морфологического анализатора // Компьютерная лингвистика и интеллектуальные технологии. 1999. Т. 2. С. 98–106; Исраилова Н.А., Бакасова П.С. Морфологический анализатор кыргызского языка // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. В 2-х томах. Т 2. – Казань: Издательство Академии наук Республики Татарстан, 2017. – 327 с.; Bird S., Klein E., Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. Beijing, 2009.; Korobov M. Morphological Analyzer and Generator for Russian and Ukrainian Languages // Analysis of Images, Social Networks and Texts: 4th International Conference, AIST 2015, Yekaterinburg, Russia, April 9–11, 2015. Communications in Computer and Information Science, Springer, 2015. P. 320–332.; Москвина А.Д., Орлова Д., Паничева П.В., Митрофанова О.А. Разработка ядра синтаксического анализатора для русского языка на основе библиотек NLTK // Компьютерная лингвистика и вычислительные онтологии. Труды XIX Международной объединенной научной конференции «Интернет и современное общество», Санкт-Петербург, 22–24 июня 2016 г. СПб, 2016. С. 44–54.; Ляшевская О.Н., Астафьева И., Бонч-Осмоловская А., Гарейшина А., Гришина Ю., Дьячков В., Ионов М., Королева А., Кудринский М., Литягина А., Лучина Е., Сидорова Е., Толдова С., Савчук С., Коваль С. Оценка методов автоматического анализа текста: морфологические парсеры русского языка // Компьютерная лингвистика и интеллектуальные технологии. 2010. № 9 (16). С. 318–326; Каневский Е.А., Боярский К.К. Морфолого-лексический анализатор

В корпусной лингвистике появилась серия исследований по вопросу решения проблемы омонимии, затрагивающая тегирования омонимических единиц и устраняющая омонимию в процессе автоматического чтения текста («снятие омонимии»). В частности, среди таких работ мы можем отметить труды Г.И.Кустовой, О.Н.Ляшевской, Е.В.Падучевой, Е.В.Рахилина, Б.П.Кобрицова, Т.И. Резниковой⁵³, Б.П. Кобрицова⁵⁴, В.В. Кукановой⁵⁵, А.А. Кретова⁵⁶, посвящённые решению этих вопросов.

В узбекском языкознании по компьютерной лингвистике и лексикографической обработке текста стоит отметить исследования М.Махмудова, М.Айимбетова, С.Каримова, Г.Джуманазаровой, А.Бабанарова, Д.Уринбаевой, А.Норова, С.Мухаммедова, Н.Абдурахмоновой⁵⁷.

Составление семантического анализатора узбекского языка не являлось предметом специального исследования. Однако в монографическом плане изучены лингвистическое обеспечение компьютерных программ на основе

и классификация текста // Прикладная лингвистика в науке и образовании. Материалы V международной научно-практической конференции. Санкт-Петербург, 2010. С. 157–163.

⁵³Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. М.: Индрик, 2005. – С. 155-174.; Ляшевская О.Н. Топологические классы имен в семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). СПб.: С.-Петербургский гос. университет, Факультет филологии и искусств, 2008. – С. 276-284.; Рахилина Е.В., Кобрицов Б.П., Кустова Г.И., Ляшевская О.Н., Шеманаева О.Ю. Многозначность как прикладная проблема: семантическая разметка в Национальном корпусе русского языка // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог-2006». – М., 2006. – С. 445-450.; Рахилина Е.В., Кустова Г.И., Ляшевская О.Н., Резникова Т.И., Шеманова О.Ю. Задачи и принципы семантической разметки лексики в НКРЯ // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы / отв. ред. В.А.Плунгян. СПб.: Не-стор-История, 2009. – С.215-239.; Ляшевская О.Н. Топологические классы имен в семантической разметке Национального корпуса русского языка // Труды международной конференции «Корпусная лингвистика-2008» (Санкт-Петербург, 6-10 октября 2008 г.). СПб.: С.-Петербургский гос. университет, Факультет филологии и искусств, 2008. – С. 276-284.

⁵⁴ Кобрицов Б.П. Модели многозначности русской предметной лексики: глобальные и локальные правила разрешения омонимии. Автореф. канд. филол. наук.– М.: РГГУ, 2004.; Кобрицов Б.П., Ляшевская О.Н., Шеманаева О.Ю. Поверхностные фильтры для разрешения семантической омонимии в текстовом корпусе // Кобозева И.М., Нариньяни А.С., Селегей В.П. (ред.), Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции Диалог'2005. – М., 2005.

⁵⁵ Куканова В.В. Принципы семантической разметки национального корпуса калмыцкого языка // http://kalmcorp.ru/sites/default/files/kukanova_25.pdf

⁵⁶ Кретов А.А. Анализ семантических помет в НКРЯ // <http://ruscorp.ru/sbornik2008/11.pdf>

⁵⁷ Махмудов М.А., Пиотровская А.А., Садыков Т. Система машинного анализа и синтеза тюркской словоформы // Переработка текста методами инженерной лингвистики. – Минск, 1982.; Мухаммедов С.А. 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. – Л., 1981.; Айимбетов М.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.; Ризаев С. Ўзбек тилининг лингвостатистик тадқиқи: Фил. фан. док. дис...автореф. – Тошкент, 2008.; Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.; Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.; Жуманазарова Г.У. Фозил Йўлдош ўгли дostonлари тилининг лингвопоэтикаси: Фил. фан. док. дис...автореф. – Тошкент, 2017.; Ўринбаева Д.Б. Ўзбек фольклори матнларининг лингвостатистик тадқиқи. – Тошкент: Фан, 2010.; Ўринбаева Д. Халқ оғзаки ижоди: жанрий-лисоний ва лингвостатистик тадқиқ муаммолари: фил. фан. док. дисс. автореф. – Самарканд, 2019.– 74 б.

глаголов действия⁵⁸, лингвистическое обеспечение машинного перевода узбекско-английского языка⁵⁹, принципы построения авторского корпуса узбекского языка⁶⁰, лингвистические модули программы по редактированию и анализу текстов на узбекском языке⁶¹, принципы построения лингвистической базы языкового корпуса⁶², лингвистические основы семантического тегирования единиц узбекского языка⁶³, лингвистические основы построения образовательного корпуса узбекского языка⁶⁴, создание базы идиом в узбекском национальном корпусе⁶⁵. Также опубликован ряд статей по проблемам построения узбекско-английского параллельного корпуса⁶⁶, автоматического анализа существительных в узбекском языке⁶⁷, формальным моделям словообразования⁶⁸, полисемии в корпусной лингвистике⁶⁹. Существуют работы, в которых освещены некоторые вопросы автоматического морфологического анализа узбекского языка⁷⁰. В работе Ш.Хамроевой в качестве объекта исследования были изучены вопросы морфоанализа узбекского языка и его лингвистического обеспечения⁷¹.

Это важно в связи с тем, что в работе были исследованы ряд проблем в области компьютерной лингвистики. Однако узбекский язык не был специально изучен программой автоматического семантического анализа –

⁵⁸ Мухаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. – Тошкент, 2006.

⁵⁹ Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида). Филол. фан. бўйича фалсафа доктори (PhD)...дис. афтореф. – Тошкент, 2018. – 52 б.

⁶⁰ Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2018. – 250 б.

⁶¹ Абжалова М.А. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар таҳрири дастури учун): Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Фарғона, 2019. – 164 б.; Абжалова М. Матнларга автолингвистик ишлов бериш тизимлари // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент, 2018. – 320 с.

⁶² Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2019. – 140 б.

⁶³ Ахмедова Д.Б. Атов бирликларини ўзбек тили корпуслари учун лексик - семантик теглашнинг лингвистик асос ва моделлари: Филол. фан. Бўйича фал. доктори (PhD) диссертацияси. – Бухоро, 2020. – 247 б.

⁶⁴ Холиёров Ў. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари. Филол. фан. бўйича фал. док-ри. дисс. афтореф. – Термиз, 2021. – Б. 16. – 52 б.

⁶⁵ Бегматова Г. Ўзбек миллий корпусида идиомалар базасини яратиш. Филол. фан. бўйича фалсафа доктори (PhD) дис.афтореф. – Термиз, 2021. – 48 б.

⁶⁶ Karimov R., Mengliev B. Theoretical fundamentals of uzbek-english parallel corpus / Journal of critical reviews. ISSN- 2394-5125. – VOL 7, ISSUE 17, 2020. – P. 73-76.; Karimov R.A., Mengliev B.R. The Role of the Parallel Corpus in Linguistics, the Importance and the Possibilities of Interpretation International Journal of engineering and Advanced Technology (IJEAT). ISSN: 2249 – 8958, Volume-8, Issue-5S3 July 2019. – P. 388-391.

⁶⁷ Орхун М. Computational analysis of uzbek nouns / Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент, 2018. – 320 с.

⁶⁸ Турсунов А. Вопросы словообразования в формальных моделях тюркских языков (на примере узбекского языка) // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». Труды конференции) – Ташкент, 2018. – 320 с.

⁶⁹ Абдурахмонова М., Рахманова А. Корпус лингвистикасида полисемия // “Компьютер лингвистикаси: муаммолар, ечим, истикболлар” Республика I илмий-техникавий конференция. – Vol. 1 №. 01 (2021).

⁷⁰ Абдурахмонова Н. Automatic morphological analyze for English-Uzbek system // Известия Кыргызский государственный технический университет им. И.Раззакова. Теоретический и прикладной научно-технический журнал. № 2 (38) – Бишкек, 2016. – С. 12-18.

⁷¹ Хамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. Филология фан. д-ри. дисс. ав-ти. – Фарғона, 2021. – 76 б.

семантическим анализатором и его лингвистической основой. В процессе подготовки диссертации перечисленные выше исследовательские работы были тщательно изучены, при необходимости обработаны и использованы в исследовании.

Соответствие исследования исследовательским планам высшего учебного заведения или исследовательского учреждения, в котором была выполнена диссертация. Диссертация выполнена в соответствии с концепцией «Лингводидактические проблемы эффективного использования возможностей узбекского языка» в Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои.

Целью диссертации является создание лингвистических основ семантического анализатора узбекского языка.

Задачи диссертации:

общее описание лингвистического анализатора, типы и его значение в информационно-поисковой системе;

освещение важности семантического анализа и синтеза в автоматическом анализе текста;

определение положения семантического анализатора в информационно-поисковой системе;

определение лингвистических основ дифференциации омонимии, полисемии и полифункциональности слов в информационно-поисковой системе;

исследование проблемы дифференциации/устранения омонимии;

разработка фильтра для устранения полифункциональных и многозначных слов в корпусе;

лингвистическое моделирование окружения омонима, полифункциональных и многозначных слов для семантического анализатора узбекского языка.

В качестве **объекта исследования** были выбраны омонимы, полифункциональные и многозначные слова, составляющие лингвистическую основу семантического анализатора узбекского языка.

Предметом исследования являются лингвистическое моделирование и систематизация омонимов, полифункциональных и многозначных слов, которые играют важную роль в создании лингвистической основы.

Методы исследования. При освещении темы исследования использовались методы описания, классификации, сравнения, статистики, моделирования.

Научная новизна исследования заключается в следующем:

выявлены морфологические, синтаксические и семантические особенности слова, проявляющиеся в процессе естественной обработки языка в информационно-поисковой системе лингвистического анализатора, в результате семантического анализа и синтеза доказано преобразование текстовых данных в понятное компьютеру лексико-морфологическое, синтаксическое и семантическое явление;

в автоматическом анализе единиц узбекского языка созданы морфологические, синтаксические лингвистические основы разграничения омонимичности, полифункциональности и многозначности, определен руководящий фактор в омонимии между существительным и глаголом – морфологический, в омонимии между прилагательным и глаголом – морфологически-синтаксический, в омонимии только между существительным и только глаголом – синтаксически-морфологический, в омонимии между только прилагательным - синтаксический, в омонимии между различными категориями морфологический, синтаксический и синтаксически-морфологический;

для полного устранения омонимии, полифункциональности и многозначности в языке не достаточно чисто семантических фильтров, поэтому на основе лингвистического фактора определены морфо-семантические, синтаксико-семантические фильтры, разработано самое необходимое средство, которое служит для различения омонимии, полифункциональности и многозначности - семантический, морфологический, морфо-семантический фильтр;

работа с семантикой слов в лингвистическом моделировании является довольно сложным процессом, для устранения путаницы и проблем при интерпретации значения определенных слов в толковых словарях, смоделированы лингвистические аспекты программного обеспечения семантического анализатора узбекского языка для омонимичных, полифункциональных и многозначных слов;

для семантического анализатора узбекского языка омонимичные слова являются основным лингвистическим обеспечением, важным вопросом является решение проблемы омонимов в корпусе, формирование базы данных омонимичных слов для снятия омонимии, для чего необходимо представить существующие в языке омонимы с их комментариями в алфавитном порядке, разработка комбинаций суффиксов для омонимов, отличающихся морфологическим путем, создание базы синтаксических фильтров для омонимов, различающихся правыми и левыми связующими сочетаниями, разработка лингвистических моделей омонимов, рекомендации по семантическим анализаторам, таким как создание математических моделей посредством лингвистических моделей.

Практические результаты исследования следующие:

доказано общее описание лингвистического анализатора, его типы и его значение в информационно-поисковой системе, роль семантического анализа и синтеза в автоматическом анализе текста;

определены такие функции, как роль семантического анализатора в информационно-поисковой системе, включая выбор правильного значения слова и связь выбранных значений со словами с независимой семантической интерпретацией;

в информационно-поисковой системе определены морфологические, синтаксические факторы дифференциации омонимии, полисемии и полифункциональности слов, созданы лингвистические основы;

были разработаны морфологические, морфо-семантические, морфо-синтаксические фильтры для того, чтобы различать/устранять омонимию, полифункциональные и многозначные слова в корпусе;

лингвистически смоделированы окружение омонимичных, полифункциональных и многозначных слов семантического анализатора для программирования.

Достоверность результатов исследования объясняется тем, что изученные материалы помогли сделать выводы, основанные на особенности узбекского языка, их обоснованности, методологической совершенности, опоре на проверенные источники при разработке лингвистических основ семантического анализатора узбекского языка.

Научная и практическая значимость результатов исследования. Исследование имеет теоретическое значение для разработки теоретических основ создания семантического анализатора узбекского языка и его лингвистических основ, в создании исследований в области компьютерной лингвистики. Практическая значимость исследования заключается в том, что компьютерная лингвистика, корпусная лингвистика и прикладное языкознание могут служить источником при составлении программ, планов и изложении тем в процессе обучения, могут быть использованы при разработке семантического анализатора узбекского языка, семантического тегирования различных типов корпуса на узбекском языке.

Внедрение результатов исследований. На основе результатов исследования значение семантического анализатора в информационно-поисковой системе, лингвистических основ разграничения омонимичных, полифункциональных и многозначных слов в информационно-поисковой системе, лингвистического моделирования окружения омонимичных, полифункциональных и многозначных слов для семантического анализатора узбекского языка были выявлены следующее:

для семантического анализатора узбекского языка омонимичные слова являются основным лингвистическим обеспечением, важным вопросом является решение проблемы омонимов в корпусе, формирование базы данных омонимичных слов для устранения омонимии, для чего необходимо представить существующие в языке омонимы с их комментариями в алфавитном порядке (1); разработка комбинаций суффиксов для омонимов, отличающихся морфологическим путем (2); создание базы синтаксических фильтров для омонимов, различающихся правыми и левыми макросочетаниями (3); разработка лингвистических моделей омонимов (4); рекомендации по семантическим анализаторам, таким как создание математических моделей посредством лингвистических моделей (5), были использованы в фундаментальном научном проекте “Функциональное словообразование в современном каракалпакском языке” FA-F1-G003, проводимом каракалпакским научно-исследовательским институтом гуманитарных наук Каракалпакского отделения Академии наук Республики Узбекистан (справочник Каракалпакского отделения Академии Наук Республики Узбекистан от 17 октября 2021 года № 17.01/265). В результате в

современном каракалпакском языке усовершенствованы модели словообразования; разработаны принципы автоматического различения омонимических, многозначных и полифункциональных значений слов в каракалпакском языке;

в информационно-поисковой системе лингвистического анализатора выявляются морфологические, синтаксические и семантические особенности слова, проявляющиеся в процессе обработки естественного языка, а в результате семантического анализа и синтеза выявляются лексико-морфологические, под которыми понимается компьютерная текстовая информация, выводы о превращении пословиц в синтаксические и семантические явления были использованы в научно-практическом проекте “каракалпакские пословицы как объект лингвистического исследования”, проводимом каракалпакским научно-исследовательским институтом гуманитарных наук Каракалпакского отделения Академии наук Республики Узбекистан (справка Каракалпакского отделения Академии Наук Республики Узбекистан от 17 октября 2021 года № 17.01/265). В результате были разработаны принципы широкого использования этих единиц в информационно-поисковой системе при определении языковой сущности каракалпакских пословиц;

в автоматическом анализе единиц узбекского языка созданы морфологические, синтаксические лингвистические основы разграничения омонимичности, полифункциональности и многозначности, морфологические в омонимии между существительным и глаголом, морфолого-синтаксические в омонимии между прилагательным и глаголом, синтаксические – в омонимии только между существительным и только глаголом, синтаксические – в омонимии только в прилагательных, морфологические – в омонимии между различными категориями., результаты работы по морфолого-синтаксическому и синтаксико-морфологическому фактору были использованы в фундаментальном научном проекте на тему “Создание электронного поэтического словаря узбекского и английского языков (образ человека, характер, природа и национальные символы)” под номером I-OT-2019-42, выполненном в Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои (приказ Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои от 23 октября 2021 года, справка № 04/1-1995). В результате исследования удалось определить место и роль омонимичных, полифункциональных и многозначных слов в описании человеческого облика, характера, природы и национальных символов, раскрыть смысловые и структурные особенности омонимичных, полифункциональных и многозначных слов в узбекском и английском языках;

на основании лингвистического фактора установлено, что для полного устранения омонимичности, полифункциональности и многозначности в языке имеют значение не только чисто семантические, но и морфосемантические, синтаксико-семантические фильтры, из выводов о

семантическом, морфологическом, морфосемантическом фильтре – важнейшем средстве, служащем для различения омонимичности, полифункциональности и многозначности использовалось 5220100 – При написании параграфа 11 “Лексика, лексико-семантический уровень и лексикология” учебника “современный узбекский язык” для студентов направления бакалавриата Филология и преподавание языков (узбекский язык) и параграфа 18 “Классификация лексики узбекского языка” (справка Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои от 20 октября 2021 года № 04/1-1975). В результате были сделаны выводы о том, что собранные в корпусе омонимичные, многозначные и полифункциональные слова и их фильтр, лингвистическая модель могут стать богатым научным источником не только лексикологии, но и лингвистической основы семантического анализатора в таких направлениях языкознания, как морфология и синтаксис;

тот факт, что работа с семантикой слов в лингвистическом моделировании является довольно сложным процессом, из выводов о лингвистических аспектах программного алгоритма, окружающего омонимичные, полифункциональные и многозначные слова для семантического анализатора узбекского языка, с целью устранения путаницы и проблем в передаче значений отдельных слов в толковом словаре, следует, что Бухарская областная телерадиокомпания «Бухара» использовала и провела семинар «Ассалом, Бухара!», «Восьмое чудо», «Литературная среда» (справка Бухарской областной телерадиокомпании от 15 октября 2021 года № 1/267). В результате были усовершенствованы выводы о формировании языковой культуры и использовании возможностей узбекского языка в современной информационной системе, что имеет важное значение для обеспечения научной обоснованности, популярности и зрительской доступности передач.

Апробация результатов исследования. Результаты данного исследования были обсуждены на 15 международных и 10 республиканских конференциях и представлены в следующих профилях: <https://orcid.org/0000-0002-2045-1912>, <https://www.researchgate.net/profile/Shahnoza-Gulyamova>, также <https://scholar.google.com/citations>.

Публикация результатов исследования. По теме диссертации опубликовано 49 научных работ, в том числе 2 монография, 17 научных статей в научных изданиях, рекомендованных Высшей аттестационной комиссией при Кабинете Министров Республики Узбекистан для публикации основных научных результатов докторских диссертаций, 5 из них опубликованы в зарубежных журналах.

Объем и структура диссертации. Диссертация состоит из введения, четырех глав, заключения, списка использованной литературы и приложения. Объем 236 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

В вводной части диссертации обоснована актуальность и необходимость темы, показана зависимость исследования от приоритетных направлений развития науки и техники республики, даны цели и задачи, описаны объект и предмет исследования, изложена научная новизна и практические результаты, раскрыта научная и практическая значимость результатов, проанализированы, апробированы, представлены сведения о структуре опубликованных работ и содержании диссертации.

Первая глава диссертации озаглавлена «**Теоретические задачи семантического анализатора**». В первом разделе главы, озаглавленной «Понятие лингвистического анализатора, его виды и характеристика», были проанализированы общие вопросы и типы лингвистического анализатора, методы анализа в мировой и узбекской компьютерной лингвистике. Пользователи используют информацию из интернет-сетей для различных целей. Как правило, понадобится точная информация, которая называется информационным объектом. Эти объекты делятся на разные типы и представляют интерес для пользователей определенных категорий. В потоке текстов на естественном языке поиск этих объектов занимает много времени, что замедляет решение конкретных задач программирования. Необходимость перевода текстов с одного языка на другой, деконструкция разных языков, компьютерные учебники, корпорация помеченных текстов, анализ и синтез текстов, создание программ аннотации и абстракции, ставит задачу устранения проблемы автоматической обработки на естественном языке.

Лингвистический анализатор – средство, представляющий морфологический, синтаксический и семантический признак слова в процессе обработки естественного языка в автоматизированной поисковой системе. Понятие анализатор как термин используется в широком смысле. В частности, наряду с лингвистикой он используется в десятках областей, таких как биология, физиология, анатомия, психология, физика, химия, для определенных целей и значений. А в лингвистике анализатор занимает важное место в автоматической обработке естественного языка. Обработка естественного языка (Natural Language Processing (NLP)) – это общее направление в области искусственного интеллекта и математической лингвистики, изучающее проблемы компьютерного анализа и синтеза текстов на естественных языках. Решение этой задачи подразумевает создание более удобной формы взаимодействия человека и компьютера. Лингвистический анализатор целесообразно подразделить на следующие виды: 1) семантический анализатор; 2) морфологический анализатор; 3) синтаксический анализатор.

Большие группы включают в себя лемматизаторы, морфоанализаторы, стеммеры, парсеры, орфокоорректоры. Каждый из этих анализаторов имеет принцип работы, цели, функцию.

Во втором разделе главы «Значение семантического анализа и синтеза в автоматическом анализе текста» рассматриваются такие вопросы, как

лингвистический анализ текста в информационно-поисковой системе, семантический анализ и синтез в автоматическом анализе, а также семантический поиск в ИКТ и его проблемы. Практическое значение имеют такие вопросы, как анализ, синтез предложения на естественном языке (SQL, SPARQL, поисковые запросы и др.), перевод предложения с одного естественного языка на другой естественный язык (система машинного перевода), задачи семантического анализа текста – TextMining, семантическая классификация текстовых ресурсов с целью формирования, понимания закономерностей запроса.

И.В.Смирнов в качестве моделей семантики указывает на следующие: 1) лексическая семантика – толкование (разъяснение) слов; 2) дистрибутивная семантика – значение, которое слова выражают посредством сочетания с другими словами; 3) модель «смысл-текст»; 4) семантика выражений; 5) модель семантических фреймов⁷².

Указанные семантические модели являются важными для функционирования семантического анализатора. Потому что семантический анализатор – это средство, который работает со значением слова. На наш взгляд, семантический анализатор узбекского языка также непосредственно работает с учетом вышеизложенного, выполняет семантический анализ. Потому что эти свойства, объединяясь, служат для наглядного представления семантического признака.

Автоматический анализ текста состоит из ряда сложных операций. В результате автоматического анализа текста его информация превращается в лексемно-морфологическое, синтаксическое и семантическое представление, понятное компьютеру. Важным является анализ и синтез автоматического текста, потому что теоретический и практический прогресс компьютерной лингвистики закладывает основу для создания эффективных систем машинного перевода, направленных на реализацию потребностей человека.

Семантический анализ является наиболее сложным направлением автоматического анализа текста. При этом между словами необходимо устанавливать семантические связи для закрепления в тексте различных языковых выражений.

В основе семантического анализа лежит утверждение, что значение слова не является элементарной смысловой единицей. Она распадается на более элементарные значения – единицы семантической лексики языка. Эти единицы семантического языка представляют собой своеобразные атомы из различных комбинаций, образующих «молекулы» из реальных слов естественного языка⁷³.

Именно семантический анализ решает проблему омонимии, которая возникает при автоматическом анализе всех языковых уровней:

1. Лексическая омонимия: совпадение звучания или правописания в словах, не имеющих элементов общего значения.

⁷²Смирнов И.В. Обработка естественного языка и искусственный интеллект // <https://cs.hse.ru/data/2017/12/06/1161306556/Slides-IntroToAI-HSE-2017-09-Smirnov.pdf>

⁷³ Апресян Ю.Д. Идеи и методы современной структурной лингвистики. М.: Просвещение, 1966. – С.254.

2. Морфологическая омонимия: суффиксальная омонимия.
3. Лексико-морфологическая омонимия (наиболее распространенный вид омонимии): совпадение двух различных лексемных словоформ.
4. Синтаксическая омонимия: синтаксическая структура, которая имеет много значений.

Модуль семантического анализа состоит из следующих двух классов: DictionaryOp и SemanticAnalysis. DictionaryOp является статическим классом, который отвечает за взаимодействие со словарем семантического анализа. Semantic-Analysis (семантический анализ) выполняет анализ принятого предложения, после чего возвращает семантическую сеть.

Семантический синтез – переход выражения от семантической к синтаксической структуре (переход от синтаксической структуры к цепочке лексико-грамматических свойств словоформ, представляющих словосочетание; переход от лексико-морфологического описания к фактической форме слова).

Поисковые системы сталкиваются с множеством проблем при реализации СП (семантического поиска). В частности, определить, что именно имел в виду пользователь при вводе поискового запроса, то есть, что слово или фраза имеют несколько значений в разных контекстах. Семантический поиск не может помочь в решении проблем, которые не могут быть решены с помощью компьютера. Семантический поиск учитывает не только контекст, но и анализ других факторов (включая текущую ситуацию, местоположение пользователя, изменение слов в семантическом поиске, синонимы, общие и специализированные запросы, «обычный язык» для запросов).

Третий раздел главы, озаглавленный «Семантический анализатор и его роль в информационно-поисковой системе» описывает вопросы изучения, цели и задачи семантического анализатора, синтаксико-семантических анализаторов, лингвистического обеспечения семантического анализатора узбекского языка. Семантический анализатор – программа для семантической обработки синтаксической структуры запроса с помощью концептуальных графиков. Концептуальный график – это семантическое или, другими словами, концептуальное представление ситуации и знаний в моделях понимания естественного языка. Узлы графика-лексических единицы, представляющие категорию и понятие, связанные между собой формальным и ассоциативным отношениями. Направленность отношений на графике всегда направлена сверху вниз – от понятий более высокого уровня обобщения к понятиям, их характеризующим. Элементарное значение определяется как смежные узлы графа, которые связаны друг с другом. Такие связи не всегда именуется, иногда в них отмечается только производная от взаимодействия двух слов (студент-институт, береза-дерево). Графа состоит из комплекса взаимосвязанных элементарных значений, входящих в допустимые сочетания, раскрывающие лексические активные и пассивные валентности того или иного слова. Связная часть графы, соединяющая два узла, образует подгруппу. Семантический анализатор является завершающим

этапом процесса обработки материала. После завершения работы лингвистического процессора анализируемая информация переносится на этап индексации.

Поскольку семантический анализатор-это система, которая проверяет правильность значения слова, его можно напрямую назвать «анализатором мышления». На наш взгляд, семантический анализатор должен обладать следующими возможностями:

во-первых, очень мощное средство проверки правильности написания (орфографии);

во-вторых, система уточнения значения слова;

в-третьих, он пытается собрать слова в обоснованные конструкции, даже если их значение немного искажено;

в-четвертых, он пытается вывести из предложения структуру, содержащую некоторые неизвестные слова или неправильно поставленные запяты.

Как видно, создание семантического анализатора – достаточно сложная задача. Для этого нам необходимо в первую очередь решить вопрос лингвистического обеспечения семантического анализатора. Семантический анализатор – это «сердце анализа»: он знает все правила, связанные с языком программирования. Семантический анализатор – это средство, служащий для различения значений существующих в языке омонимических, полифункциональных и многозначных слов.

II глава диссертации озаглавлена «**Лингвистические основы дифференциации омонимов в информационно-поисковой системе**». В первом разделе главы «Традиционное языкознание и исследование омонимии в компьютерной лингвистике» рассматриваются такие вопросы, как история изучения феномена омонимии в мировом языкознании, изучение феномена омонимии в тюркологии и изучение феномена омонимии в узбекском языкознании.

С.Ульманн назвал омонимию «лингвистической универсальностью». Он утверждает, что омонимия не является абсолютной универсальной, но что в мире можно найти язык, для которого омонимия не существует, без омонимии «такой язык будет более эффективным средством общения». С.Ульманн считал, что омонимия – это статистическая универсальность, то есть относительность. Также ученый считал, что омонимы существуют во многих языках мира [Ульманн, 1970].

В мировой лингвистике теории, касающиеся изучения феномена омонимии, можно сгруппировать следующим образом: 1. Феномен омонимии существующая в языке многозначная реальность. 2. Явление, не имеющее значения для языка, «отвлекающее». Следовательно, исходя из противоположных представлений об омонимах, можно сказать, что омоним – это слово с одинаковым произношением, одинаковым внешним видом, разным содержанием.

Омонимия стал предметом исследований таких ученых-тюркологов, как К.Аханова (казахский язык) [1956], М.Миртожиева (узбекский язык) [1964],

А.Х.Ахтямова (башкирский язык) [1966], Р.Бекджановой (киргизский язык) [1967], Б.С.Салимгареевой (татарский язык) [1972], А.Гельдимуратова (туркменский язык) [1974], А.Меметова (крымско-татарский язык) [1977], А.А.Гасанова (азербайджанский язык) [1980]. Эти ученые-тюркологи опираются на теорию омонимии, разработанную А.И.Смирницким, В.В.Виноградовым, О.С.Ахмановой и другими учеными-лингвистами. Помимо всего этого, проблема омонимов в тюркологии анализируется в научных трудах Н.А.Баскакова, Н.К.Дмитрия. Также проводили исследования ряд таких российских и узбекских ученых, как В.В.Радлов, Э.В.Севортян, А.А.Юлдашева, К.К.Юдахина, М.Х.Ахтямова, С.И.Ибрагимов.

Омонимы в каждом языке являются историческим явлением лексикологии, они являются не показателем бедности языка, а показателем богатства и разнообразия лексического фонда.

Второй раздел главы, озаглавленный «Вопросы дифференциации/устранения омонимии в корпусе», представляет собой обзор исследования омонимов в компьютерной лингвистике и опыт их семантического анализа, методы анализа омонимов, формирование базы данных омонимических слов для семантического анализатора узбекского языка. Как известно, изучение, исследование и анализ языковых явлений имеют свое значение во все времена. В мировом языкознании такие явления, как омонимия, синонимия, антонимия и полисемия, изучаются с точки зрения компьютерной лингвистики, которая является новой для узбекской лингвистики, ее практические результаты используются в таких целях, как составление и совершенствование корпуса. В то время как на ранних и средних этапах языкознания изучалось возникновение, развитие и отношение феномена омонимии к смежным явлениям, в современной лингвистике на повестку дня встал вопрос о его выделении в корпус. Для выполнения этой задачи существующие словари служат информационными банками.

В корпусной лингвистике возник ряд исследований, посвященных решению проблемы омонимии, вопросу о тегировании омонимических единиц и устранению омонимии в процессе автоматического чтения текста («снятие омонимии»). В частности, Г.И.Кустова, О.Н.Ляшевская, Э.В.Падучева, Э.В.Рахилина, Б.П.Кобрицов, Т.И.Резникова⁷⁴,

⁷⁴Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Семантическая разметка лексикологии в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. М.: 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.). СПб.: С.-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». – М., 2006. – С. 445-450.; Рахилина Е.В., Кустова Г.И., Ляшевская О.Н., Резникова Т.И., Шеманова О.Ю. Задачи и принципы семантической разметки лексикологии НКРЯ // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы / отв. ред. В.А.Плунгян. СПб.: Не-стор-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.). СПб.: С.-Петербургский гос. университет, Факультет филологии и искусств, 2008. – С. 276-284.

Б.П.Кобрицов⁷⁵, В.В.Куканова⁷⁶, А.А.Кретов⁷⁷ опубликовали ряд работ, посвященных решению этих вопросов.

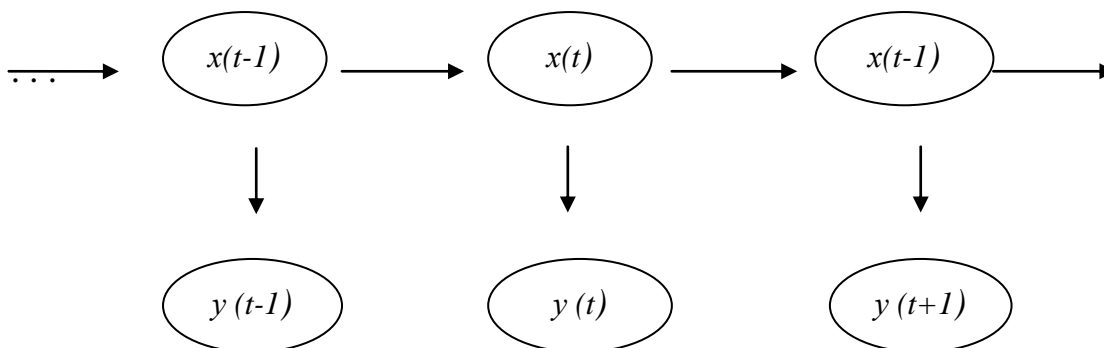
Снятие омонимии актуально во многих программах компьютерной лингвистики, в частности, в поисковых системах. Потому что это может повысить точность обработки определенных классов запросов или уменьшить объем хранимой информации. В целях сравнения снятия омонимии в англоязычных текстах были выбраны два метода: метод базовых векторов и скрытая Марковская модель (СММ). Эти приемы дают возможность частично снять омонимию для слов, отсутствующих в словаре. Известная скрытая модель Маркова (Hidden Markov Model (HMM) tagging) направлена на определение омонимов на основе вероятностного подхода. Основная идея алгоритма заключается в том, что для каждого слова, включенного в предложение, необходимо выбрать грамматический класс (тег), который максимизирует следующую функцию:

$$P(\text{word}/\text{tag}) * P(\text{tag}/\text{previous } n \text{ tags})$$

$P(\text{tag}/\text{previous } n \text{ tags})$ – условная вероятность того, что данный тег будет сформирован после определения предыдущего n тегов.

$P(\text{word}/\text{tag})$ – условная вероятность того, что слово word , сгенерированное в заданном месте, имеет тег данного грамматического класса.

Скрытая Марковская модель – статистическая модель, имитирующая работу процесса с неизвестными параметрами. Решение неизвестных параметров на основе переменных наблюдений представляется задачей⁷⁸.



1 - картина. Общая структура СММ.

В этом случае становится ясно, что значение скрытой переменной $x(t)$ зависит только от значения скрытой переменной $x(t-1)$. Эта связь (подчинение) называется «особенностью Маркова». В то же время значение наблюдаемой переменной $y(t)$ будет зависеть только от значения $x(t)$

⁷⁵ Кобрицов Б.П. Модели многозначности русской предметной лексики: глобальные и локальные правила разрешения омонимии. Автореф. канд. филол. наук.-М.: РГГУ, 2004.; Кобрицов Б.П., Ляшевская О.Н., Шеманаева О.Ю. Поверхностные фильтры для разрешения семантической омонимии в текстовом корпусе // Кобозева И.М., Нариньяни А.С., Селегей В.П. (ред.), Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции Диалог'2005.-М., 2005.

⁷⁶КукановаВ.В. Принципы семантической разметки национального корпуса калмыцкого языка // http://kalmcorp.org/sites/default/files/kukanova_25.pdf

⁷⁷ Кретов А.А. Анализ семантических помет в НКРЯ // <http://ruscorp.org.ru/sbornik2008/11.pdf>

⁷⁸Ветров Д. Скрытые марковские модели // http://www.machinelearning.ru/wikiimages/8/83/GM12_3.pdf.

скрытой переменной. Для анализа слов, которых нет в словаре, используются трехбуквенные постфиксы. В результате решения задачи определения морфологических признаков сводится к вычислению последовательности наиболее подходящих значений скрытых значений:

$$y'(t) = \operatorname{argmax} P(y|x)$$

Алгоритм СММ имеет очень высокую вычислительную сложность (реализован классическим методом динамического программирования Витерби, аналогом которого является Левенштейн, используемый для вычисления расстояния), и на практике, как правило, для его уменьшения используются различные упрощающие предположения (например, не более трех последовательностей слов).

В третьем разделе главы, озаглавленном «Факторы дифференциации омонимии» обсуждаются морфологические факторы дифференциации омонимии, феномен разговорной речи и омонимии, синтаксические факторы дифференциации омонимии. Как известно, языковые единицы отличаются друг от друга рядом факторов. В частности, к ним относятся лексические, морфологические, синтаксические факторы. При дифференциации единиц каждый фактор будет иметь свое место и значение. При их дифференциации наблюдается, что один фактор является ведущим, другой фактор может быть ослаблен или дифференцируется на основе сотрудничества факторов. Дифференциация омонимов, являющихся объектом исследования, осуществляется по ряду путей. Как отмечалось ранее, для того, чтобы семантический анализатор мог различать омонимы, необходимо изначально разработать их лингвистический фильтр и модели. Это требует лингвистических знаний, подхода. Уже сейчас обработка естественного языка (NLP) является одним из наиболее важных и актуальных направлений синтеза искусственного интеллекта и математической лингвистики. Здесь основное действие направлено на то, чтобы человек мог анализировать текст, выраженный на естественном языке, и находить и совершенствовать способность компьютера «воспринимать», обрабатывать эту информацию. Обработка естественного языка – это задача, связанная с искусственным интеллектом, поскольку она включает в себя высокий уровень понимания окружающего мира и человеческих эмоций. Очевидно, что решение этих задач требует от исследователя достаточных знаний в области лингвистики и искусственного интеллекта.

Для исследования лингвистической базой послужили «Толковый словарь омонимов» Ш. Рахматуллаева и «Толковый словарь узбекского языка» (5 томов). Основываясь на их категории, мы сгруппировали их следующим образом: 1) омонимы в рамках одной категории слов; 2) омонимы, наблюдаемые между существительным // глаголом, а также прилагательным // категорией глаголов; 3) омонимы, образованные в рамках разных частей речи.

Слова-омонимы, которые встречаются в рамках существительного // глагола и прилагательного // глагола, различаются в основном на основе

морфологического фактора. Это происходит с помощью ряда морфологических средств.

В дифференциации слов-омонимов играет важную роль морфологические и синтаксические факторы. В омонимии между существительным и глаголом морфологический; между прилагательным и глаголом морфолого-синтаксический; только существительного и только в кругу глагола синтаксическо-морфологический; только внутри круга ряда прилагательных синтаксический; в разных частях речи руководствует морфологический, синтаксический, морфолого-синтаксический и синтаксическо-морфологический фактор.

III глава исследования озаглавлена «**Лингвистические основы дифференциации полифункциональных и многозначных слов в информационно-поисковой системе**». Первый раздел озаглавлен «Исследование полифункциональных и многозначных слов в традиционной лингвистике».

Многозначность занимает значительное место в любом языке. Богатство языка измеряется не только словосочетаниями, пословицами, но и лексическими (словарными) значениями слов. Многообразие значений слов, то есть многозначность слов – полисемия, имеет свое место в богатстве языка.

Случай многозначности в языке очень давно находится в центре внимания ученых. Известный лексикограф Махмуд Кашгари в своей книге «Девону луғатит турк» отмечает тюркские многозначные слова. Великий мыслитель Алишер Навои в своем труде «Муҳокамат ул-луғатайн» также уделяет особое внимание значениям слов. В произведениях Навои почти все слова многозначны, поэтому лексикографы всегда сталкиваются с полисемантизмом при толковании значения слова. В результате в истории узбекской лексики сложилась своя последовательная система толкования значения слова. Во всех словарях, созданных в этот период, толкование значения слова велось по принципу «от общего к частному», сначала указывалось общее, правильное значение, затем – переносное значения слова⁷⁹.

Оценивание многозначности вызвала противоречивые мнения среди лингвистов. Некоторые ученые считают, что в идеальном языке слово должно иметь только одно значение, и каждое значение должно иметь отдельное название. Но это может показаться удобным на первый взгляд, ведь однообразие слов снижает возможности языка, лишая его национальной самобытности. Большинство ученых по праву считают многозначность слов не слабостью языка, а проявлением богатства языка. Многозначность признает безграничные возможности языка, поскольку словарный запас – это не только слова, но и способность слов приобретать новые смысловые различия во всем многообразии их значений. Освоение новых значений в

⁷⁹Нурмонов А. Ўзбек тилшунослиги тарихи. – Тошкент: Ўзбекистон, 2002. – Б. 48. – 232 б.

слове дает простор для творческого использования лексических ресурсов языка.

Если каждое лексическое значение многозначного слова рассматривать как одно кольцо, то они соединяются между собой цепочкой. Выпадение одного из них будет разрывом цепочки, т.е. лексических значений многозначного слова. Ученый М.Миртожиев особо подчеркивает образование омонимов на основе того же процесса за счет многозначного слова.

Это не оправдано тем фактом, что некоторые лингвисты вообще не признают феномена полисемии, рассматривая его как «небольшое изменение значения слова делает его другим словом». Напротив, многозначные слова как синонимы, антонимы и омонимы наглядно демонстрируют развитие языка, его богатство, способность усваивать в одном слове много значений, принципы экономии в языке. Отдельно следует выделить еще один аспект, когда выборочное использование целесообразного значения многозначного слова, использование его на месте требует от носителя языка остроумия. Создание программы дифференциации многозначных слов для национального корпуса узбекского языка имеет важное практическое значение для дифференциации значений слов в национальном корпусе узбекского языка.

Полифункциональность связана с явлением конверсии. О.С.Ахманова в «Словаре лингвистических терминов» определяет «конверсию как явление прихода определенного слова на место другого слова без получения какого-либо аффикса»⁸⁰. Понятие полифункциональности в лингвистике необычайно разнообразно. Полифункциональность выделяется как свойство единиц языковых уровней: исследователи придерживаются мнения о полифункциональности морфемы, слова (словоформы), словосочетаний. Иногда полифункциональность понимается более узко, иногда очень широко.

Полифункциональность также может быть в пределах одного грамматического класса. Например, среди вспомогательных слов распространена полифункциональность. Особенно это характерно для союзов и частиц. Смысловой единицей таких слов доказано М.И. Черемисиной: «В определенном положении союз принимает на себя частицу, и наоборот, в определенных положениях частица принимает на себя функцию союза»⁸¹.

В нашем языкознании проведен ряд работ по вопросу полифункциональности. Анализу этой проблеме посвящены исследования Ё.Тожиева, Ж.Элтазарова, Н.Кодировой, Г.Сулаймоновой, Т.Валиева, А.Ботировой и они представляют особую ценность⁸². По вопросу о полифункциональности языковых единиц высказываются различные точки

⁸⁰ Ахманова О.С. Словарь лингвистических терминов. – Москва, 1966. – С. 197. – 607 с.

⁸¹ Черемисина М.И. Сравнительные конструкции русского языка. – Москва, 1976. – 270 с.

⁸² Кодирова Н.А. Ўзбек тилидаги аффиксларда полифункционаллик: Филол. фан. номз. дисс. автореф. – Тошкент, 2002. – 26 б. – Б.16; Сулаймонова Г.А. Ўзбек тилида от туркумидаги сўзларнинг полифункционаллиги: Филол. фан. номз. дисс. автореф. – Тошкент, 2010. – 26 б. – Б. 8.; Элтазаров Ж. Д. Ўзбек тилида сўз туркумларининг ўзаро алоқаси ва кўчиши: Филол. фан. докт. дисс. автореф. – Тошкент, 2008. – Б. 26. – 42 б.

зрения. Обобщая их, можно разделить на следующие группы: 1) изучение полифункциональности, присущей языковым единицам, в рамках полисемии; 2) союзное звено между полисемией и омонимией; 3) синкретизм значений, присущих различным частям речи в семантической структуре слова; 4) единица с ослабленным свойством многозначности, но не достигшая степени омонимичности.

Раздел данной главы, озаглавленный «Из опыта фильтрации полифункциональных и многозначных слов для снятия многозначности в корпусе», описывает исследование полифункциональных слов в компьютерной лингвистике, рассматривается вопрос о фильтрации многозначных слов и их представлении в корпусе языка, об опыте снятия многозначности. В ходе исследования мы не наблюдали ни одного исследования, посвященного изучению полифункциональности в компьютерной лингвистике, автоматической реализации дифференциации значений слов. Объясняется это, на наш взгляд, тем, что в теоретической лингвистике нет единого научного заключения о явлении. В некоторых небольших исследованиях, освещающих проблему передачи полисемии в корпусе, к этому явлению была высказана попутная реакция. В том числе в статье И.М.Кобозевой «Полисемия дискурсивных слов и возможностей и разрешений в контексте предложения (на примере слова *вот*)» приводятся некоторые рассуждения о том, что слово *вот* является полифункциональным словом как по функции частицы, так и по функции местоимения и выражается в русском национальном корпусе⁸³. Наряду с полифункциональным в работах некоторых лингвистов встречается термин «многофункциональность». В том числе А.Лефевре говорит: «Лексический элемент, который выполняет более одной грамматической функции, является многофункциональным». В качестве синонимов используются термины полифункциональный и многофункциональный⁸⁴.

Основываясь на исследованиях, проведенных в традиционном направлении узбекского языкознания по полифункциональности, о значении локализации в фильтрации таких слов, была предпринята попытка лингвистически решить проблему полифункциональных слов для семантического анализатора, освещая вопросы лингвистических факторов в семантической фильтрации полифункциональных слов в тексте.

В корпусной лингвистике возник ряд исследований, посвященных решению проблемы полисемантического слова, тегированию полисемантической единицы, устранению полисемии в процессе автоматического чтения текста («снятие полисемии»)⁸⁵. При решении задачи

⁸³ Кобозева И.М. Полисемия дискурсивных слов и возможности ее разрешения в контексте предложения (на примере слова *Вот*) // Труды международной конференции «Диалог 2007». – С. 250-255.

⁸⁴ Alan Reed Libert *Hypersynonymy for Polyfunctionality* //

https://www.researchgate.net/publication/329881879_Hypersynonymy_for_Polyfunctionality

⁸⁵ Шеманаева О.Ю., Кустова Г.И., Ляшевская О.Н., Рахилина Е.В. Семантические фильтры для разрешения многозначности в Национальном корпусе русского языка: прилагательные // Иомдин Л.Л., Лауфер Н.И., Нариньяни А.С., Селегей В.П. (ред.). Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2007». – 2007. – С. 582-587.; Толдова С.Ю., Кустова Г.И., Ляшевская О.Н. Семантические фильтры для разрешения многозначности в Национальном корпусе русского языка: глаголы // Компьютерная лингвистика и

лексической многозначности необходимо: 1. Определение значения каждого слова, связанного с текстом. 2. Выбор наиболее подходящего значения на основе слова, встречающегося в контексте.

Значение многозначного слова в корпусе отличается не цифрой, как в простом толковом словаре, а семантическим признаком, например: пилить (пень) – «физическое движение», пилить (мужа) – «речь». Программы используют семантические фильтры. Если в словаре каждый знак ставится в соответствие с соответствующим значением, то в корпусных текстах он ставится перед словом целиком, поэтому при автоматическом расположении знака невозможно выделить значение слова. Программа многозначности использует семантические фильтры. Многозначное слово используется в предложении в определенном смысле (исключая игру слов), что означает, что значение слова соответствует контексту.

В третьем разделе «Факторы дифференциации полифункциональных и многозначных слов внутри себя», рассматриваются такие вопросы, как явление локализации при различии полифункциональных и многозначных слов, лингвистические факторы дифференциации полифункциональных слов, факторы дифференциации многозначных слов. В узбекском языкознании существует исследование, посвященное выражению локальности на лексическом уровне⁸⁶, в котором основное внимание уделяется раскрытию локальных семантик в лексемах и системообразующей (интегративной) функции этих семантик в лексической системе языка. В узбекском языке соотношение локум и локализация освещена в диссертации М.В.Эргашова⁸⁷. В диссертации Н.Шириновой показано явление локализации как релевантно значимых синтаксических приемов в разграничении предметных и признака-атрибутивных значений в лексике. Широко применяемым для всех языков фактором усиления признака, в определенной степени смягчающего предметность, в языкознании является локализация (соположение), в котором обычно подчиненный компонент практически сливается со значением подлежащего, реализуя значение признака-атрибута. Этот метод очень широко используется в нашем языке: каменный мост, золотые часы...⁸⁸ Синтаксическая связь состоит, во-первых, из взаимосвязей элементов синтаксической единицы, то есть служит для выражения синтаксических отношений между словами. Во-вторых, создает синтаксическую структуру предложения и словосочетания. В-третьих, создает условия для реализации лексического значения слова. Словоформы, вспомогательные слова, лексические элементы, интонация, порядок слов являются средствами,

интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4-8 июня 2008 г.). – Вып. 7 (14). – М.: РГГУ, 2008.; Рахилина Е.В., Ляшевская О.Н., Кобрицов Б.П., Кустова Г.И., Шеманаева О.Ю. Многозначность как прикладная проблема: Лексико-семантическая разметка в Национальном корпусе русского языка // Лауфер Н.И., Нариньяни А.С., Селегей В.П. (ред.). Компьютерная лингвистика и интеллектуальные технологии: Труды между народной конференции “Диалог 2006”, – 2006. – С. 445-450.

⁸⁶ Курбонова Б. Локалик ва унинг ўзбек тилида ифодаланиши: Филол. фанлари номзоди дисс. – Тошкент, 2007. – 107 б.

⁸⁷ Эргашова М. Ўзбек тилида локум ва локализация муносабати. Филол. фанлари номзоди дисс. ав-ти. – Тошкент, 2011. – 21 б.

⁸⁸ Ширинова Н. Ўзбек тилида белги-хусусият ва предметлик маъноларининг фарқлинишида градуонимик қаторлар. Филол. фанлари номзоди дисс. – Тошкент, 2010. – 160 б. – Б. 85.

производящими синтаксическую связь. В лексике необходимость отличать синтаксический фактор локализации (размещения) от простого синтаксического примыкания проанализировано в диссертации С.Назаровой⁸⁹. Понятие локализации важно для определения многозначных и полифункциональных значений слова в соответствии с расположением словоформ. Потому что, как слово определенного значения диктует определенную синтаксическую позицию, так и определенная синтаксическая позиция диктует слово того значения, которое соответствует этой позиции. Другими словами, чем больше семантической природы имеет синтаксис, тем больше семантики, в свою очередь, имеет синтаксическая природа⁹⁰.

Поскольку создатель языка – это человек, не следует упускать из виду взаимодействие языковых и прагматических факторов для реализации языковых средств. При анализе полифункциональных слов, являющихся объектом исследования, наряду с лингвистическими факторами следует учитывать и значение речевой ситуации, речевого окружения.

В «Толковом словаре омонимов узбекского языка» такие глаголы, как *дай, смотри, приходи, получи, отправь, останься*, данные как омонимы, уместно рассматривать как полифункциональные слова. Теоретические основы этого были проанализированы в первой главе данной главы. Остановимся на факторах разграничения полифункциональности между самостоятельным и вспомогательным глаголом:

Бер (дай) – глагол. Держать в руках, отдавать в распоряжение, обеспечивать.

Бер (дай) – вспомогательный глагол. К лексическому значению, которое обозначает самостоятельный глагол: 1) «выполнение для кого-то» добавляет грамматическое значение. 2) добавляет грамматическое значение «длительность, повторяемость»⁹¹.

<i>Бер</i> самостоятельный глагол (доминирующая часть в СС))	
китобни, меҳрни, бахтни	бер
менга, унга, бизга	
уйда, кўчада, маҳаллада, ишда	
пулдан, ширинликдан, чойдан	
<i>Бер</i> самостоятельный глагол (подчиненная часть в СС))	
бериб беришга берса берганда (-гунча, -гач) бергандан кейин	кел
бериб берса	юбор

⁸⁹ Назарова С. Бирикмаларда сўзларнинг эркин боғланиш омиллари. Филол. фанлари номзоди дисс. – Тошкент, 1997. – 20 с.

⁹⁰ Гак В. Г. К проблеме семантической синтагматики. «Проблемы структурной лингвистики». – М., 1971. – С. 367.

⁹¹ Рахматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. – Тошкент: Ўқитувчи, 1984. – Б. 29. – 214 б.

берганда (-гунча, -гач) бергандан кейин	
бериб беришни (-га, -да) берса берганда (-гунча, -гач) бергандан кейин	айт
<i>Бер</i> мустақил феъл (КФСҚнинг етакчи қисми) <i>Дай</i> самостоятельный глагол (доминирующая часть в СС)	
бериб (-а)	кўр
	юбор
	сол

Через предлоги глагол *бер (дай)* реализуется как самостоятельное слово. Будучи самостоятельным глаголом, он выполнял функцию повелительного слова, направляя согласные формы. В примерах *бер кел, бериб юбор «бериб»* был самостоятельным глаголом и подчиненным компонентом в прымикании.

<i>Бер</i> вспомогательный глагол	
жеркиб	бер
айтиб	
куйлаб	
текшириб	
Айта	
кўра	
топшира	

Здесь вспомогательный глагол является в функции вспомогательной части словосочетания. Самостоятельная и вспомогательная часть связаны через формы наречий. Таким образом, когда данный глагол является самостоятельным, он действует как самостоятельный и подчиненный. А когда это вспомогательный глагол, он связан только с самостоятельным глаголом, что означает действие, выполненное для кого-то.

Из анализа было замечено, что значение многозначных слов проявляется во взаимодействии морфологических и синтаксических факторов. Выделение полисемантических слов из «Толкового словаря омонимов узбекского языка», подбор коннотаций на основе языковых факторов и создание совершенной базы послужат важным ресурсом для будущего семантического анализатора.

IV глава диссертации озаглавлена «Вопросы лингвистического моделирования окружения омонимичных, полифункциональных и многозначных слов для семантического анализатора узбекского языка». В разделе «Лингвистическое моделирование окружения омонимичных слов» анализируются проблемы лингвистического моделирования окружения

омонимичных слов, наблюдаемых в рамках одной категории, существительных/глаголов, прилагательных/глаголов, а также омонимичных слов, возникающих в рамках разных категорий. В компьютерной лингвистике важное значение имеет лингвистическое моделирование. На основе созданной модели создается программное обеспечение, решается вопрос о представлении языковой единицы в корпусе. Для семантического анализатора узбекского языка важным является вопрос лингвистического моделирования окружения омонимичных слов. Исходя из категориального характера каждого из видов омонимов, необходимо построить их языковое обеспечение и через него языковую модель. Потому что окружение омонимичных слов в рамках одного созвездия слов не может быть сделано как моделирование окружения омонимичных слов, которые встречаются в пределах существительного и глагола, а также прилагательных и глаголов или омонимичных слов, которые возникают в рамках другого созвездия. Ведь омонимичные слова в рамках одной категории невозможно определить по грамматическим формам. Как известно, омонимичные слова, встречающиеся в пределах одной категории, сохраняют свою омонимичность даже при сложении грамматических форм. Например:

Асорат I от. Туткунлик, куллик, эркислик ҳолати.

(Осложнение I существительное. Состояние неволи, рабства, безбрачия).

Асорат II от. Бўлиб ўтганнинг қолдиғи, излари.

(Осложнение I существительное. Пережиток прошлого, следы⁹²).

Асорат + нинг/ни/га/да/дан = асоратнинг/ни/га/да/дан

(Осложнение + нинг/ни/га/да/дан = нинг/ни/га/да/дан осложнения)

Асорат + нинг/ни/га/да/дан = асоратнинг/ни/га/да/дан

(Осложнение + нинг/ни/га/да/дан = нинг/ни/га/да/дан осложнения)

Видно, что через такой фильтр невозможно различить омонимы, встречающиеся в пределах одной категории. Поэтому такие слова потребовало создания особенные базы фильтров.

В «Толковом словаре омонимов узбекского языка» и «Толковом словаре узбекского языка», послуживших лингвистической базой, мы определили количество и вид омонимов в пределах одной категории слов следующим образом:

1) омонимичные слова только в пределах категории существительного – 1060;

2) омонимичные слова только в пределах категории прилагательных – 82;

3) омонимичные слова только в пределах глагола – 235.

Снять омонимию в узбекском языке можно следующими способами:

1) сочетание с предыдущим самостоятельным словом – W_P ;

2) сочетание со следующим самостоятельным словом – W_F .

⁹² Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. Тошкент: Ўқитувчи, 1984. – 108 б. – Б 22.

В лингвистическом моделировании окружения омонимичных слов мы опирались на соединение частных значений, потому что слово – это многогранное явление вч и одно слово может иметь много сочетаний. Речевое употребление слова также может «тянуть» за собой прилагательные. Работа с семантикой слова в данном случае остается сложной, особенно для корпусной лингвистики ее моделирования. Желательно ограничить этот охват частным значением. Это облегчает создание базы семантических фильтров и моделирование омонимов.

Семантическая фильтрация омонимичных слов, встречающихся в пределах одной категории, может осуществляться следующими способами:

1. Определение омонимов только в пределах рода существительных по тому, какое слово следует за ним, принимая падежные. Это можно смоделировать следующим образом:

$H_N + CS_aff = V$. Здесь: $H_N = \{\text{существительное-омоним}\}$;

$CS_aff = \{-, \text{-нинг}, \text{-ни}, \text{-га}, \text{-да}, \text{-дан}\}$; $V = \{\text{глагольное слово}\}$.

$H_N + CS_aff = N/Num/Adj/Adv/Pr/Mim$.

Здесь: $H_N = \{\text{существительное-омоним}\}$;

$CS_aff = \{-, \text{-нинг}, \text{-ни}, \text{-га}, \text{-да}, \text{-дан}\}$; $N = \{\text{существительное}\}$,
 $Num = \{\text{существительное}\}$, $Adj = \{\text{прилагательное}\}$, $Adv = \{\text{наречие}\}$,
 $Pr = \{\text{местоимение}\}$, $Mim = \{\text{подражание}\}$.

$Adj + H_N$. Где: $Adj = \{\text{прилагательное}\}$, $H_N = \{\text{существительное-омоним}\}$. Существительные-омонимы, работающие по этой модели, тоже присутствуют в лексиконе, они сочетаются с прилагательным, которое предшествует им, определяющим значение существительного-омонима.

Это инвариантные модели, предназначенные только для определения омонимов в пределах категории существительных, что реализуется в ряде подмоделей.

Определение омонимов, встречающихся только в пределах прилагательного словосочетания, с помощью специального метода, основанного на сочетаемости частных значений, которые могут идти после него.

$H_{adj} + N$. Здесь: $H_{adj} = \{\text{прилагательное-омоним}\}$; $N = \{\text{существительное слово}\}$;

$H_{adj} + V$. Здесь: $H_{adj} = \{\text{прилагательное-омоним}\}$; $V = \{\text{глагол-слово}\}$.

Только дифференциация омонимов в категориях глаголов осуществляется следующими способами:

1) глагол-омоним 2-го лица определяется на основе того, что перед ним закрепляются частные значения, если они находятся в склонении повелительного наклонения или неопределенной формы глагола (-моу).

2) если глагол-омоним образован с помощью суффиксов (деепричастия, причастия, категория состояния -ш, -иш, -в, -ув), то глагол-омоним образует соединение со словом, которое стоит либо перед ним, либо после него.

Они могут быть смоделированы следующим образом:

$WG + H_V$. Здесь: $WG = \{\text{существительное, наречие, прилагательное, числительное, местоимение}\}$; $H_V = \{\text{глагол-омоним}\}$

$Y_r^f + O_f^n$. Здесь: $Y_r^f = \{\text{ведущий глагол: -б/-б, -а/-й}\}$; $O_f^n = \{\text{глагол-омоним}\}$.

Это инвариантные модели, предназначенные только для определения омонимов в рамках группы глагольных слов, которые реализуются в ряде подмоделей.

Мы выделили из лингвистической базы 120 омонимичных слов, встречающихся в пределах существительного и глагола, а также прилагательного и глагола. Из этих омонимичных слов 102 являются существительными и глаголами, а 18 прилагательными и глаголами.

Ари I от. Парда қанотли чақадиган хашарот.

Ари I (оса) существительное. Перпончатокрылое жалающее насекомое.

Ари II феъл. Йўқолмоқ, тугамоқ, кетмоқ⁹³.

Ари II глагол. Исчезать, заканчиваться, уходить.

1) ари+ни/лар/ча =существительное.

2) ари+ма/ган/гунча/гин/май/ди=глагол

Ая I от. Фарзандни дунёга келтирган аёл, она.

Ая I существительное. Женщина, родившая ребенка, мать.

Ая II феъл. Авайламоқ, эҳтиёт қилмоқ⁹⁴.

Ая II глагол. Беречь.

1) ая+жон/м/нг/си/миз/нгиз/нинг/ни/га/да/дан/лар=существительное.

2) ая+ ма/май/мас/ган/р/б/гач/гунча/гудек/й/нг/син/нгиз = глагол.

Мы различали омонимы, подобные приведенным выше, между существительным и глаголом через такой фильтр. Существительное-омоним можно определить с помощью словообразовательных и синтаксических формообразующих суффиксов, присущих существительному. Это можно смоделировать следующим образом:

$N_N + LF_N + SF_N = N$

Здесь N_N – омонимичное слово, LF_N (Lexical forms of the noun) – лексические формы существительного, SF_N (Syntactic forms of the noun) – синтаксические формы существительного, N (Noun) – словосочетание существительного.

А глагол-омоним можно определить с помощью словообразовательных и синтаксических формообразующих аффиксов, присущих глаголу. Смоделируем это следующим образом:

$N_V + LF_V + SF_V = V$

Здесь N_V – омонимичное слово, LF_V (Lexical forms of the verb) – лексические формы глагола, SF_V (Syntactic forms of the verb) – синтаксические формы глагола, V (Verb) – словосочетание глагола. Эти модели реализуются в ряде субмоделей.

В «Толковом словаре омонимов узбекского языка», послужившем лингвистической базой исследования, омонимичные слова, возникающие в рамках различных категорий, разнообразны, и лингвистическое

⁹³ Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли луғати. Тошкент: Ўқитувчи, 1984. – 108 б. – Б 21.

⁹⁴ Ўша манба. – Б. 24.

моделирование их окружения потребовало разработки индивидуального фильтра, потому что к разным категориям относятся также два, три, четыре категорий, и можно наблюдать, что некоторые омонимы имеют до 2 и даже 5-6 значений. Из классификации различных категорий омонимов видно, что для различения омонимичных слов в рамках разных категорий требуется фильтрующая база в соответствии со значением каждого слова.

Второй раздел главы озаглавлен «Лингвистическое моделирование окружения полифункциональных слов». Вопрос о полифункциональном слове не нашел полного решения в мировой лингвистике, и мы не встречали ни одного исследования его автоматического определения, снятия в корпусе, языкового фильтра и модели. В этом разделе была предпринята попытка обобщить существующие теоретические взгляды на узбекском языке и с их помощью провести предварительные работы по моделированию полифункциональных слов для семантического анализатора.

Ряд полифункциональных слов между существительных-прилагательных, такие как *золото, серебро, чугу́н, больной, железо, дерево (бревно), хромой, нижний, выпуклый*, моделируются следующим образом:

$P_i^N + V$. Здесь: $P_i^N = \{\text{полифункциональное слово, которое было существительным, } i=1..n\}$; $V = \{\text{глагол}\}$.

$P_i^N + V = \{\text{найти/добыть/купить/дать золото...}\}$

$P_1^{ADJ} + N$. Здесь: $P_{ADJ} = \{\text{полифункциональное слово, которое было прилагательным, } i=1..n\}$; $N = \{\text{от}\}$

$P_1^{ADJ} + N = \{\text{золотое кольцо/браслет/часы/голова/рука...}\}$

Третий раздел главы озаглавлен «Лингвистическое моделирование окружения многозначных слов». Как известно, снятие лексической многозначности (англ. *word sense disambiguation, WSD*) – нерешенная проблема обработки естественного языка, заключающаяся в задаче выбора значения многозначного слова или фразы. Эта проблема возникает при дискурсивном анализе, оптимизации релевантности результатов поисковыми системами, решении анафорических ссылок, изучении лингвистических связей текста, анализе умозаключений. При различении многозначности также опирается на принцип конъюнкции. Значение слова, определяется словом, которое предшествует или следует за ним. Учитывая внутренние возможности и особенности узбекского языка, можно выделить следующие способы моделирования многозначных слов (на примере слова *ёқа*):

$N_j^{ps.1} + PS_1$

$N_j^{ps.1} = \text{многозначные слова существительного в сочетании с начальным (собственным) значением слова.}$

$N_j^{ps.1} = \{\text{рубашка, одежда,...}\}; j = 1..n$

$PS_1 = \text{многозначное слово, пришедшее в первоначальном (собственном, денотативном) значении.}$

Например: *воротник рубашки, воротник платья*

$PS_1 + V_j^{ps.}$

$V_j^{ps.} = \text{многозначные слова глагольного рода, сочетающиеся с первоначальным (собственным) значением слова.}$

$V_j^{ps.} = \{\text{шить, расстегивать, расстегивать,...}\}; j = 1 \dots n$

Например: расстегнуть воротник, сшить воротник

$Adj_j^{ps.} + PS_1$

$Adj_j^{ps.}$ = слова из категории прилагательных, сочетающиеся с первоначальным (собственным) значением многозначного слова.

$Adj_j^{ps.} = \{\text{широкий, узкий, длинный, короткий ...}\}; j = 1 \dots n$

Например: *широкий воротник, узкий воротник*

Полисемантика моделируется следующим образом, когда слово используется в другом (переносном, коннотативном) значении:

$N_j^{ps.2} + PS_2$

$N_j^{ps.2} + PS_2$ = слова из категории существительных с переносными значениями многозначного слова.

$N_j^{ps.2} = \{\text{канавы, озеро, смерть, канал, улица, дорога,...}\}; j = 1 \dots n$

PS_2 = выражает использование слова в переносном смысле.

Например: *вдоль канавы, вдоль озера, вдоль канала, вдоль улицы, вдоль дороги*

Моделирование является основным орудием обработки естественного языка⁹⁵. Для семантического анализатора существует ряд проблем в моделировании омонимичных, многозначных и полифункциональных слов. Во-первых, работа со семантикой слова – процесс достаточно сложный. Потому что слово-это многогранное явление. Во-вторых, в толковом словаре узбекского языка есть путаница в толковании значений некоторых слов, что еще больше запутывает вопрос. Для создания совершенной языковой базы для семантического анализатора требуется индивидуальная работа над значением каждого слова с использованием моделей, построенных выше. Для этого значения слов, их толкование, категория, какой характер они носят, должны быть внесены в хранилище данных в идеальном состоянии. Семантический анализатор служит для определения значения слова на основе предложенных в работе форм, для показа массива примеров.

ЗАКЛЮЧЕНИЕ

1. Лингвистический анализатор – средство, представляющий морфологический, синтаксический и семантический признак слова в процессе обработки естественного языка в автоматизированной поисковой системе. К таким средствам относятся лемматизатор, морфоанализатор, стеммер, парсер, орфокоорректоры. Каждый из этих анализаторов имеет принцип работы, назначение, функцию.

2. Для разработки семантического анализатора важно сначала осветить значение семантического анализа и синтеза в автоматическом анализе текста. Автоматический семантический анализ является одной из актуальных и в то же время сложных задач компьютерной лингвистики. Он является

⁹⁵Хамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. Монография. – GlobeEdit, 2020. – 253 б. – Б. 197.

необходимым этапом создания любой системы, связанной с человеческим интеллектом.

3. В мировой лингвистике проводились обширные исследования морфологического и синтаксического анализатора, но специальной практической работы по семантическому анализатору не проводилось. Потому что формализация семантики-чрезвычайно сложная задача. Принципиальная сложность проблемы заключается в том, что при изучении содержания (значения) приходится выходить за пределы языка и обращаться к внешнему миру, к классификации предметов и представлений, лежащих вне сферы языка.

Автоматический семантический анализ и синтез являются одной из сложных задач компьютерной лингвистики и имеют большое значение в автоматическом анализе текста. Совершенствование процесса семантического анализа и синтеза в информационно-поисковой системе, семантического поиска и решения его задач открывает путь к уточнению задачи семантического анализатора, к построению в дальнейшем системы (семантического анализатора). Для функционирования системы требуется, чтобы изначально в ИПС (информационно-поисковой системе) был систематизирован процесс смыслового анализа и синтеза. От значения семантического анализа и синтеза в автоматическом анализе текста напрямую зависит создание языкового обеспечения узбекского языка, определение семантического анализатора и его места в информационно-поисковой системе.

4. Решение проблемы устранения омонимии требует тщательного изучения опыта исследования омонимичных слов в компьютерной лингвистике и их семантического анализа, методов анализа омонимов, их целесообразного практического использования и классификации омонимов по категориям. В работе омонимы слова были разделены на такие группы, как омонимы в пределах одной категории слов (1); омонимы, наблюдаемые между существительным и глаголом, а также между прилагательным и глаголом(2); омонимы, возникающие в рамках разных категорий (3). Именно такая классификация имеет лингвистическое обоснование. Потому что для создания фильтра омонимов важны роль и значение факторов, которые их дифференцируют.

5. Языковые единицы отличаются друг от друга лексическими, морфологическими, синтаксическими факторами. При дифференциации единиц у каждого фактора есть своя функция. Когда один фактор является ведущим, может наблюдаться ослабление другого фактора или дифференциация факторов на основе сотрудничества (когда несколько факторов влияют вместе). В результате исследования было установлено, что морфологический и синтаксический фактор играет важную роль в дифференциации омонимичного слова. В омонимии между существительным и глаголом руководствует морфологический; между прилагательным и глаголом морфолого-синтаксический; между только существительного и только в кругу глагола синтаксико-морфологический; только внутри круга

ряда прилагательных синтаксических; в разных частях речи руководствует морфологический, синтаксический, морфолого-синтаксический и синтаксическо-морфологический фактор.

6. Для формирования базы данных омонимичных слов для семантического анализатора узбекского языка необходимо (1) собрать и классифицировать в алфавитном порядке имеющиеся в языке омонимические слова из «Толковом словаре омонимов» и «Толковом словаре узбекского языка» (в 5 томах) с комментариями на основе категорий; (2) разработать комбинации суффиксов для омонимов, различающихся морфологическим путем; (3) создать базу синтаксических фильтров для омонимов, различающихся правыми и левыми суффиксами; (4) на основе разработка языковых моделей омонимов; (5) были разработаны рекомендации, как создание математических моделей с помощью лингвистических моделей.

7. В различении значений полифункциональных и многозначных слов работает метод локализации. Морфологические и синтаксические факторы играют важную роль в дифференциации этих явлений. Проявлялось также значение роли морфологических форм, основанных на принципе соединения.

При различении значения омонимов важным фактором послужил случай коллокации. С помощью коллокационных словосочетаний можно выяснить значение омонимов и определить частоту их применения. Омоним теряет свою омонимичность в составе коллокации. Потому что в составе сочетания реализуется только одно значение. С помощью коллокации можно получить статистику о том, сколько слов и словосочетаний используется. С его помощью корпусная лингвистика может определить сильные и слабые соединения.

8. Устранение омонимичных, многозначных и полифункциональных слов актуально во многих программах компьютерной лингвистики, в частности в поисковых системах. Потому что это может повысить точность обработки определенных классов запросов или уменьшить объем хранимой информации. Из анализа видно, что лидирует скрытая Марковская модель снятия омонимических, многозначных и полифункциональных слов. Эта модель важна тем, что она направлена на определение омонимов на основе вероятностного подхода.

9. В ходе исследования было выявлено 7 инвариантных моделей и 6 подмоделей для определения омонимов в рамках одной категории; омонимичные слова, встречающиеся в пределах существительных и глаголов, а также прилагательных и глаголов, - 4 инвариантных и 2 подмодели; а окружение омонимичных слов, возникающих в рамках разных категорий, - 20 инвариантных моделей, 3 подмодели. Разработано 12 моделей и 37 подмоделей для определения некоторых полифункциональных слов в узбекском языке. Для определения некоторых многозначных слов составлено 10 моделей.

10. Из анализа видно, что для семантического анализатора существует ряд проблем в моделировании омонимических, полисемантических и полифункциональных слов. Потому что слово – это

явление многогранное, а работа со значением – сложный процесс. В толковом словаре узбекского языка существует путаница в значении некоторых слов, что еще больше усложняет вопрос. Для создания совершенной лингвистической базы семантического анализатора необходима индивидуальная работа над значением каждого слова. Для этого значения слов, толкование, категории слов, какой бы характер они ни носили, должны быть включены в хранилище данных в идеальном состоянии.

**SCIENTIFIC COUNCIL FOR AWARDING SCIENTIFIC DEGREES
DSc.03/30.12.2019.Fil.05.02 AT FERGANA STATE UNIVERSITY**

**TASHKENT STATE UNIVERSITY OF UZBEK LANGUAGE AND
LITERATURE**

GULYAMOVA SHAKHNOZA KAKHRAMONOVNA

**LINGUISTIC FOUNDATIONS OF THE SEMANTIC ANALYZER OF THE
UZBEK LANGUAGE**

10.00.11 – Theory of language, applied and computational linguistics

**LINGUISTIC FOUNDATIONS OF THE SEMANTIC ANALYZER OF THE
UZBEK LANGUAGE**

ABSTRACT
of the doctoral thesis (DSc) on philological sciences

Fergana – 2022

The theme of DSc dissertation is registered by the Supreme Attestation Commission under the Cabinet of Ministers of the Republic of Uzbekistan under № B2021.2.DSc/FI325

The dissertation was performed at the Tashkent State University of the Uzbek Language and Literature named after Alisher Navoi.

The dissertation abstract in three languages (Uzbek, Russian and English (resume)) has been placed on the website (www.fdu.uz) of Scientific Council and on the website of "ZiyoNet" information-educational portal (www.ziynet.uz).

Scientific consultant:

Mengliyev Bakhtiyor Rajabovich
Doctor of Philology, professor

Official opponents:

Samsitova Luiza Khamzinovna
Doctor of Philology, professor

Uluqov Nosirjon Muhammadallyevich
Doctor of Philology, professor

Shaxabitdinova Shokhida Khashimovna
Doctor of Philology, professor

Leading organization:

Karshi State University

Defense of the dissertation will take place on 1 » March 2022, at 10⁰⁰ at a meeting of Scientific Council DS.c.03/30.12.2019.Fil.05.02 under Fergana State University (Address: 19, Murabbiylar Street, Fergana, 100151. Tel: (+99873) 244-44-29; fax: (+99873) 244-66-03; e-mail: info@fdu.uz.

The dissertation is available in the Information Resource center of Fergana State University (registered under № 50). Address: 19, Murabbiylar Str., Fergana, 100151. Tel: (+99873) 244-44-02.

The abstract of the dissertation is distributed on « 8 » February 2022
(Mailing report № 44 dated « 9 » February 2022).

Sh. Iskandarova
Chairman of the Doctor's Degree
awarding Scientific Council, Doctor of
Science, Professor

M. Zokirov
Scientific secretary of the
Scientific Council awarding scientific
degrees, Candidate of Philological
Sciences

A. Mannajanov
Chairman of the Scientific
Seminar at the Doctor's Degree awarding
Scientific Council, Doctor of Science,
Professor

INTRODUCTION (abstract of DSc dissertation)

The purpose of the dissertation is to create the linguistic foundations of the semantic analyzer of the Uzbek language.

The object of the research is the linguistic support of the semantical analyzer of the Uzbek language.

The subject of research is linguistic modeling and systematization of homonyms, polyfunctional and polysemous words, which play an important role in creating a linguistic foundation.

Task of the research consists of the following:

general description of the linguistic analyzer, types and its meaning in the information retrieval system;

highlighting the importance of semantic analysis and synthesis in automatic text analysis;

determination of the position of the semantic analyzer in the information retrieval system;

determination of the linguistic foundations of differentiation of homonymy, polysemy and polyfunctionality of words in an information retrieval system;

study of the problem of differentiation / elimination of homonymy;

development of a filter to eliminate polyfunctional and ambiguous words in the corpus;

linguistic modeling of the environment of a homonym, polyfunctional and polysemous words for the semantic analyzer of the Uzbek language.

Scientific novelty of the research consists of the following:

it has been determined that the linguistic analyzer represents the morphological, syntactic, and semantic features of a word that are manifested in the process of processing natural language in an information retrieval system and as a result of semantic analysis and synthesis, text data has been proven to become a lexical-morphological, syntactic, and semantic phenomenon that a computer can understand;

morphological, syntactic linguistic bases of differentiation of homonymy, polyfunctionality and polysemy in the automatic analysis of Uzbek language units are developed. Morphological in the homonymy between noun and verb, morphological-syntactic in the homonymy between adjective and verb, syntactic-morphological only in the homonymy of noun and verb, syntactic only in the context of adjective homonymy, and morphological, syntactic, morphological-syntactic and syntactic-morphological factors in the homonymy of different part of speeches are determined;

linguistic factors have shown that pure semantic filters are not enough to completely eliminate homonymy, polyfunctionality and ambiguity in language, and that morpho-semantic, syntactic-semantic filters are also important. Therefore, the most necessary tool, namely semantic, morphological, morpho-semantic filter to distinguishing homonymy, polyfunctionality and ambiguity has been developed;

working with word semantics in linguistic modeling is a more complex process, the linguistic aspects of homonymous, polyfunctional and polysemous word-surround software for the Uzbek semantic analyzer are modeled in order to eliminate confusion and problems in the meaning of some words in the explanatory dictionary;

for the semantic analyzer of the Uzbek language and for the basic linguistic support of homonymous words, to solve the problem of homonyms in the corpus, to eliminate homonymy, to form a database of homonymous words, it is necessary to explain the homonyms in the language in alphabetical order as well as recommendations have also been made on semantic analyzers such as the development of a combination of suffixes for morphologically differentiated homonyms, the creation of a syntactic filter base for homonyms with former and latter combinations, the development of linguistic models of homonyms based on the created filter base, and the creation of mathematical models through linguistic models.

Scientific and practical significance of the research results.

The research plays a theoretical role in the development of the theoretical basis of the semantic analyzer of the Uzbek language and its linguistic basis, in the creation of research in the field of computer linguistics. The practical significance of the research is explained by the fact that computer linguistics, corpus linguistics and applied linguistics serve as a source in the teaching of programs, plans and descriptions of topics, can be used in the development of semantic analyzer of Uzbek language, semantic tagging of different types of corpus in Uzbek.

The structure and scope of the dissertation. The dissertation consists of an introduction, four chapters, conclusion and the list of references. The volume of the study is 236 pages.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS
I бўлим (I часть; I part)

1. Гулямова Ш.К. Ўзбек тили семантик анализаторининг лингвистик асослари. Монография. – Германия: GlobeEdit, 2021. – 276 б.
2. Гулямова Ш.К. Корпус лингвистикаси – замонавий амалий тилшуносликнинг устувор йўналиши // Бухоро давлат университети илмий ахбороти. – Бухоро, 2020. – № 4. – Б. 106-111. (10.00.00; № 1)
3. Гулямова Ш.К. Жаҳон тилшунослигида омонимия ҳодисасининг ўрганилиш тарихи ва босқичлари // Сўз санъати халқаро журнали. 1-сон, 3-жилд. – Тошкент, 2020. – Б. 104-109. (10.00.00; № 31)
4. Гулямова Ш.К. Кўп маъноли сўз ва унинг ўрганилиши хусусида // Фан ва жамият. Илмий-методик журнал. – № 3. 2020. – Б. 44-46. (10.00.00; № 2)
5. Гулямова Ш.К. The problem of polyfunctionality in the traditional and computational linguistics // Middle European scientific bulletin. ISSN 2694-9970. Journal Impact Factor = 5.985. – 104-107 p. ((23) Scientific Journal Impact Factor; (14) ResearchBib)
6. Гулямова Ш.К. Semantic analysis and synthesis in the automatic analysis of the text // Бухоро давлат университети илмий ахбороти. – Бухоро, 2021. – № 1. – Б. 112-125. (10.00.00; № 1)
7. Гулямова Ш.К. Linguistic analyzer and its types // Academicia: An International Multidisciplinary Research Journal (IF: 7, 492). Vol. 11, Issue 2, February 2021. ((23) Scientific Journal Impact Factor)
8. Mengliev B., Shahabitdinova Sh., Khamroeva Sh., Gulyamova Sh., Botirova A. The Morphological Analysis and Synthesis of Word Forms in the Linguistic Analyzer // Linguistica antverpiensia, 2021 Issue-1 www.hivt.be ISSN: 0304-2294. – P. 703-712. ((3) SCOPUS)
9. Khamroeva Sh.M., Gulyamova Sh.K. Electronic dictionaries – the product of applied linguistics // ISJ Theoretical & Applied Science, 07 (87), 2020, July, 463-466. – SoI: <http://s-o-i.org/1.1/TAS-07-87-88>.((17) Open Academic Journals Index; (35) CrossRef)
10. Хамроева Ш.М., Гулямова Ш.К. Туркий корпус лингвистикаси ривожланиши ва бугунги ҳолати // Сўз санъати халқаро журнали. – 2020. №1 – Б. 227-235. – DOI <http://dx.doi.org/10.26739/2181-9297-2020-1> (10.00.00; № 31)
11. Гулямова Ш.К. Ахборот-қидирув тизимида семантик анализ ва синтез масаласи. – ЎзМУ хабарлари. – Тошкент, 2021. – № 1. – Б. 227-231. (10.00.00; № 15)
12. Гулямова Ш.К. Семантик анализатор ва унинг ахборот-қидирув тизимида тутган ўрни. – “Тил, таълим, таржима” халқаро журнали. – Самарқанд, 2021. – № 2. – Б. 30-41. (10.00.00; №10)
13. Гулямова Ш.К., Ахмедова Х.И. Омоним сўзларнинг лингвистик асоси, математик модели ва алгоритмлари (От ва феъл ҳамда сифат ва феъл туркуми

доирасида). – Қўқон ДПИ. Илмий хабарлар. – Қўқон, 2021. – № 2. – Б. 103-110.

14. Гулямова Ш.К. Семантик анализатор учун омонимликни фарқлашнинг морфологик ва синтактик омиллари // ЎзМУ хабарлари. – Тошкент, 2021. – № 1/5/1. – Б. 239-243. (10.00.00; № 15)

15. Гулямова Ш.К. Бир туркум доирасидаги омоним сўзлар қуршовини лингвистик моделлаштириш масаласи // Фан ва жамият. Илмий-методик журнал. – № 3. 2021. – Б. 67-69. (10.00.00; № 2)

16. Гулямова Ш., Ҳамроева Ш. Туркий тилли корпус тузишнинг назарий ва амалий масалалари хусусида // “Тамаддун нури” журнали. – № 4. 2021. – Б. 43-47. (10.00.00; № 28)

17. Гулямова Ш.К., Ахмедова Х.И. Ўзбек тили семантик анализатори учун омоним сўзларлар маълумотлар базасини шакллантириш масаласи хусусида // Сўз санъати халқаро журнали. 4-жилд, 3-сон. – Тошкент, 2021. – Б. 326-334. – Doi 10.26739/2181-9297 (10.00.00; № 31)

18. Gulyamova Sh. From the Experience of Filtering Polysemy Words and Their Use in the Language Corpus // Pindus Journal Of Culture, Literature, and ELT. ISSN: 2792 – 1883 Volume 9. <https://literature.academicjournal.io> ((23) Scientific Journal Impact Factor; (14) ResearchBib)

19. Гулямова Ш.К. Корпус лингвистикасида коллокация ҳодисаси // International conference science and education. – Turkiya, 2021. – Б. 117-119.

20. Гулямова Ш.К. Relation of polysemy and homonymy // Дистанционные возможности и достижения науки. Международная научно-практическая конференция. – Киев, Украина, 2020. Август. – Р. 94-96.

21. Гулямова Ш.К. Дунё ва ўзбек компьютер тилшунослигида анализатор тадқиқи // Замонавий таълимда рақамли технологиялар: Филология ва педагогика соҳасида замонавий тенденциялар ва ривожланиш омиллари. Халқаро илмий-амалий масофавий конференция материаллари. – Тошкент, 2021, февраль. – Б. 52-55.

22. Гулямова Ш.К. Анъанавий тилшунослик ва компьютер лингвистикасида полифункционаллик масаласи // Давлат тили – ижтимоий тараққиёт ва миллий юксалиш мезони. Республика илмий-амалий конференция материаллари. – Бухоро, 2020, октябрь. – Б. 366-369.

23. Гулямова Ш.К. Семантик анализатор учун омонимликни фарқлаш омилларининг айрим масалалари хусусида // “Ўзбекистоннинг умидли ёшлари” мавзусидаги 6-сон Республика илмий талабалар, магистрлар ёш тадқиқотчилар ва мустақил изланувчилар учун онлайн конференция материаллари тўплами. – Тошкент: Тадқиқот, 2021. – Б. 74-76.

II бўлим (II часть; II part)

24. Менглиев Б., Элов Б., Гулямова Ш., Ахмедова Х. От ва феъл, сифат ва феъл сўз туркумлари орасидаги омонимликни фарқловчи лингвистик, математик моделлар ва воситарнинг маълумотлар базаси // – № DGU 12071 (муаллифлик гувоҳномаси). 11.08.2021.

25. Гулямова Ш., Ахмедова Х. Бир сўз туркуми доирасидаги омонимликни фарқловчи лингвистик, математик моделлар ва воситаларнинг маълумотлар базаси // – № DGU 13109 (муаллифлик гувоҳномаси). 18.11.2021.

26. Гулямова Ш.К. Ўзбек тили семантик анализатори учун лингвистик таъминот масаласи. Монография. – Германия: GlobeEdit, 2020. – 80 б.

27. Гулямова Ш.К. Корпус лингвистикасига кириш. Амалий топшириқлар учун қўлланма. – Германия: GlobeEdit, 2020. – 40 б.

28. Гулямова Ш.К. История изучения феномена омонимии в узбекском языкознании // International Scientific Journal “Internauka”. <https://doi.org/10.25313/2520-2057-2020-14>

29. Гулямова Ш.К. История изучения феномена омонимии в тюркологии языкознании // “Интернаука” научный журнал. № 32 (161). 2020. Сентябрь. – С. 45-48.

30. Гулямова Ш.К. Полисемия лингвистик ходиса сифатида // “Ўзбекистонда илмий-амалий тадқиқотлар” мавзусидаги онлайн конференция. 31-июль. № 18. – Тошкент, 2020. – Б. 26-28.

31. Хамроева Ш., Гулямова Ш. Роль электронных словарей в развитии компьютерной лингвистики // Инновационное развитие науки и образования. Международная научно-практическая конференция. – Казахстан, 2020, август. – Б. 88-93.

32. Хамроева Ш., Гулямова Ш. Электрон луғатлар – компьютер лексикографияси маҳсули // Давлат тили – ижтимоий тараққиёт ва миллий юксалиш мезони. Республика илмий-амалий конференция материаллари. – Бухоро, 2020. – Б. 459-465.

33. Гулямова Ш.К. Полифункционаллик хусусида // “Ўзбекистонда илмий-амалий тадқиқотлар” мавзусидаги онлайн конференция. 15-август. № 19. – Тошкент, 2020. – Б. 46-48.

34. Гулямова Ш.К. Полисемантик сўзларнинг тил корпусида берилиши масаласи хусусида // Замонавий таълимда рақамли технологиялар: филология ва педагогика соҳасида замонавий тенденциялар ва ривожланиш омиллари. Халқаро илмий-амалий масофавий конференция. – Тошкент, 2020, сентябрь. – Б. 167-169.

35. Гулямова Ш.К. Омонимларнинг компьютер лингвистикасидаги тадқиқи // Ўзбек тилини дунё миқёсида кенг тарғиб қилиш бўйича ҳамкорлик истиқболлари. Халқаро илмий-амалий анжуман материаллари. – Тошкент, 2020, октябрь. – Б. 401-408.

36. Гулямова Ш.К. Семантик-синтактик анализаторлар хусусида // Компьютер лингвистикаси: муаммолар, ечим, истиқболлар. Республика I илмий-техникавий конференция. Vol. 1 №. 01 (2021). – Б. 110-115.

37. Гулямова Ш.К. Лингвистик анализатор хусусида // Компьютер лингвистикаси: муаммо ва ечимлар. Халқаро илмий-амалий масофавий конференция. – Тошкент, 2021, апрель. – Б. 104-111.

38. Гулямова Ш.К. Корпус лингвистикасида омонимияни филтрлаш масаласи (от ва феъл ҳамда сифат ва феъл туркуми доирасида) // Ўзбек

миллий ва таълимий корпусларини яратишнинг назарий ва амалий масалалари. Халқаро илмий-амалий конференция. – Тошкент, 2021. – Б. 105-110.

39. Гулямова Ш.К., Холова Ш.Ш. Семантик анализатор борасидаги изланишлар хусусида // Ўзбек миллий ва таълимий корпусларини яратишнинг назарий ва амалий масалалари. Халқаро илмий-амалий конференция. – Тошкент, 2021. – Б. 265-270.

40. Гулямова Ш.К. Компьютер лингвистикасида моделлаштириш масаласи // Замонавий филология илми ва лингводидактика масалалари. Республика илмий-амалий конференция материаллари тўплами. – Қарши, 2021. – В. 13-14.

41. Гулямова Ш.К., Ҳамроева Ш. Туркий тилли корпуслар типологияси // Integration into the world and connection of sciences. International scientific and practical online conference. – Baku, Azerbaijan. June, 2021. – Б. 61-66.

42. Ҳамроева Ш., Гулямова Ш. Туркий тилли корпус тузиш тажрибасидан // Uluslararası türk bilim öğrenci kurultayı. – Nur-Sultan, 2021. – Б. 303-319.

43. Гулямова Ш.К., Ҳамроева Ш.М. Туркий тилли корпуслар типологияси // Uluslararası türk bilim öğrenci kurultayı. – Nur-Sultan, 2021. – Б. 419-425.

44. Гулямова Ш.К. Ўзбек тили семантик анализатори учун омонимияни филтрлаш масаласи хусусида // “Ўзбекистоннинг умидли ёшлари” мавзусидаги 7-сон Республика илмий талабалар, магистрлар ёш тадқиқотчилар ва мустақил изланувчилар учун онлайн конференция материаллари тўплами. – Тошкент: Тадқиқот, 2021. – Б. 6-10.

45. Гулямова Ш. Полисемантик сўзларни моделлаштириш масаласи // Ўзбек тили тараққиёти ва халқаро ҳамкорлик масалалари. Халқаро конференция материаллари тўплами. – Тошкент, 2021. – Б. 243-246.

46. Менглиев Б., Гулямова Ш. Корпус лингвистикасида кўп маънолилиқни бартараф этиш тажрибаси хусусида // Ўзбек амалий тилшунослиги ва лингводидактика масалалари. Республика илмий-амалий конференция материаллари тўплами. – Тошкент, 2021. – Б. 9-14.

47. Гулямова Ш. Полифункционал ва кўп маъноли сўзларни фарқлашда локализация ҳодисаси // Ўзбек амалий тилшунослиги ва лингводидактика масалалари. Республика илмий-амалий конференция материаллари тўплами. – Тошкент, 2021. – Б. 27-29.

48. Гулямова Ш. Кўп маъноли сўзлар қуршовини лингвистик моделлаштириш // Филологик тадқиқотлар: муаммо ва ечим. Халқаро илмий-назарий конференция материаллари тўплами. – Бухоро, 2021. – Б. 122-124.

49. Менглиев Б., Гулямова Ш. Вопрос лингвистического моделирования омонимичных слов в рамках одной категории в узбекском языке // Туркий тилларни компьютерда қайта ишлаш – TurkLang-2021. VIII халқаро конференция материаллари тўплами.

Автореферат Фарғона давлат унивеерситети “FarDU. ILMIY
XABARLAR – Scientific journal of the Fergana State University” журнали
таҳририятида таҳрирдан ўтказилди.

Босишга рухсат этилди: 2022 й. Нашриёт босма табоғи: – 3,375.

Шартли босма табоғи: – 1,6875. Бичими 84x108 1/16. Адади 100.

Баҳоси келишилган нархда.

“Poligraf Super Servis” МЧЖ

150114, Фарғона вилояти, Фарғона шаҳар, Авиасозлар кўчаси, 2-уй.

