

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

ХАМРОЕВА ШАҲЛО МИРДЖОНОВА

ЎЗБЕК ТИЛИ МОРФОЛОГИК АНАЛИЗАТОРИНИНГ ЛИНГВИСТИК
ТАЪМИНОТИ

10.00.11 – Тил назарияси, амалий ва компьютер лингвистикаси

ФИЛОЛОГИЯ ФАНЛАРИ ДОКТОРИ (DSc) ДИССЕРТАЦИЯСИ
АВТОРЕФЕРАТИ

Фарғона – 2021

Докторлик (DSc) диссертацияси автореферати мундарижаси

Оглавление автореферата докторской (DSc) диссертации

Contents of the abstract of doctoral (DSc) dissertation

Хамроева Шахло Мирджоновна

Ўзбек тили морфологик анализаторининг лингвистик таъминоти..... 3

Хамроева Шахло Мирджоновна

**Лингвистическое обеспечение морфологического анализатора
узбекского языка..... 35**

Khamroeva Shahlo Mirjonovna

Linguistic support of the morphological analyzer of the Uzbek language... 70

Эълон қилинган ишлар рўйхати

Список опубликованных работ

List of published works..... 74

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

ХАМРОЕВА ШАҲЛО МИРДЖОНОВА

ЎЗБЕК ТИЛИ МОРФОЛОГИК АНАЛИЗАТОРИНИНГ ЛИНГВИСТИК
ТАЪМИНОТИ

10.00.11 – Тил назарияси, амалий ва компьютер лингвистикаси

ФИЛОЛОГИЯ ФАНЛАРИ ДОКТОРИ (DSc) ДИССЕРТАЦИЯСИ
АВТОРЕФЕРАТИ

Фарғона – 2021

Фан доктори (DSc) диссертацияси маълуми Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида В2020.4.DSc/Fil227 рақам билан рўйхатга олинган.

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

Диссертация автореферати уч тилда (ўзбек, инглиз, рус (резюме)) Илмий кенгашнинг вебсаҳифасида (www.fdu.uz) ва «ZiyoNet» Ахборот таълим порталида (www.ziynet.uz) жойлантирилган.

Илмий маслаҳатчи:

Менглиев Бахтиёр Ражабович
филология фанлари доктори, профессор

Расмий оппонентлар:

Самситова Луиза Хамзиевна
филология фанлари доктори, профессор

Искандарова Шарифа Мадалпиевна
филология фанлари доктори, профессор

Шахабиддинова Шехида Хашимовна
филология фанлари доктори, профессор

Етакчи ташкилот:

Самарқанд давлат университети

Диссертация ҳимояси Фарғона давлат университети ҳузуридаги филология фанлари бўйича илмий даражалар берувчи DSc.03/30.12.2019.Fil.05.02 рақамли Илмий кенгашнинг 2021 йил «28» июль соат 9:00 даги мажлисида бўлиб ўтди (Манзил: 100151, Фарғона шаҳри, Мураббийлар кўчаси, 19-уй. Тел.: (99873) 244-66-02; факс: (99873) 244-44-02; e-mail: info@fdu.uz).

Диссертация билан Фарғона давлат университетининг Ахборот-ресурс марказида таништиш мумкин (1/2 рақами билан рўйхатга олинган). Манзил: 100151, Фарғона шаҳри, Мураббийлар кўчаси, 19-уй. Тел.: (373) 244-44-02; ; e-mail: info@fdu.uz.

Диссертация автореферати 2021 йил «17» июль кунин тарқатилди. (2021 йил «17» июль даги 22 рақамли реестр баённомаси)



А.А.Қосимов

Илмий даражалар берувчи илмий
кенгаш раиси, филол.ф.д., профессор

М.Т.Зокиров

Илмий даражалар берувчи илмий
кенгаш аъзоси, филол.ф.н., доцент

А.Б.Мамнжонов

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

КИРИШ (докторлик (DSc) диссертацияси аннотацияси)

Тадқиқот мавзусининг долзарблиги ва зарурати. Жаҳон компьютер лингвистикасида автоматик морфологик таҳлил XX асрнинг 50-йилларидан табиий тилга ишлов бериш, математик лингвистика ва машина таржимаси амалиётида қўллана бошланди. Туркий тилларнинг морфологик анализаторларини ишлаб чиқишнинг амалий асосларига ўтган асрнинг 60-йилларидан эътибор қаратилди. Фандаги бу янгилаишлар ахборот технологияларини тилшуносликка татбиқ этиш билан боғлиқ истиқболли илмий йўналишларнинг вужудга келишига асос бўлди. Зеро, бу ахборот-қидирув тизими, корпус лингвистикаси ва машина таржимасида морфологик таҳлилнинг умумий тамойилларини ўрганиш, таҳлил тизимлари лингвистик таъминоти учун формал тил ва грамматикани яратиш тизимини ишлаб чиқишнинг заруриятини юзага келтирди. Бу эса сунъий интеллектнинг автоматик таржима, компьютер таҳлили ва таҳрири, тезаурус, электрон луғат борасидаги имкониятларининг амалий аҳамиятини белгилайди.

Дунё тилшунослигида XX аср охирида автоматик таржима сифатини яхшилаш, тилни лингвистик моделлаштириш, муайян тиллар доирасида сўзларни леммалаш назарияси, алгоритмининг яратиш каби масалалар кун тартибига қўйилди. Натижада ахборот-қидирув тизимида фаол қўлланувчи стеммер, орфографик корректор (имло муҳаррири) сингари иловалар ишлаб чиқилди. Бинобарин, тил корпусларини яратиш, уларни лингвистик аннотациялаш, матнни автоматик қайта ишлайдиган дастурлар – морфоанализатор, лемматизатор, стеммер, парсер, орфокорректорларни ишлаб чиқиш компьютер лингвистикасининг илмий-назарий муаммоларига айланди.

Ўзбек компьютер лингвистикаси соҳасида тил корпуси, нутқ синтезатори, автоматик таржима, сунъий интеллект, ўзбек тилини тушуниш ва қайта ишлаш борасида илмий изланишларга эътибор қаратилмоқда. Лекин миллий тил морфологик анализаторининг дастурини тузиш масаласи илмий-назарий планда ўрганилмаган. Зеро, “...давлат тилининг софлигини сақлаш, уни бойитиб бориш ва аҳолининг нутқ маданиятини ошириш; давлат тилининг замонавий ахборот технологиялари ва коммуникацияларига фаол интеграциялашувини таъминлаш”¹ ҳозирги кунда ўзбек компьютер лингвистикаси олдида турган долзарб вазифалардан биридир. Шунга кўра, ўзбек тилига замонавий ахборот технологиялари воситасида ишлов берилишига эришиш, бунинг учун тил корпуслари, электрон таржимон, тезаурус, орфокорректор каби дастлабки автоматик ишлов воситалари ҳамда уларнинг лингвистик таъминотини яратиш сингари масалалар муҳим аҳамият касб этади.

Ўзбекистон Республикаси Президентининг 2016 йил 13 майдаги ПФ-4997-сон “Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти

¹Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг 2020 йил 202 октябрдаги «Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида»ги ПФ-6084-сон фармони // <https://lex.uz/docs/5058351>

университетини ташкил этиш тўғрисида”, 2017 йил 7 февралдаги ПФ-4947-сон “Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида”, 2019 йил 21 октябрдаги ПФ-5850-сон “Ўзбек тилининг давлат тили сифатидаги нуфузи ва мавқеини тубдан ошириш чора-тадбирлари тўғрисида”, 2020 йил 20 октябрдаги ПФ-6084-сон “Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида”ги фармонлари, 2017 йил 17 февралдаги ПҚ-2789-сон “Фанлар академияси фаолияти, илмий-тадқиқот ишларини ташкил этиш, бошқариш ва молиялаштиришни янада такомиллаштириш чора-тадбирлари тўғрисида”, 2019 йил 4 октябрдаги ПҚ-4479-сон “Ўзбекистон Республикасининг “Давлат тили ҳақида”ги Қонуни қабул қилинганлигининг ўттиз йиллигини кенг нишонлаш тўғрисида”ги қарорлари, 2020 йил 24 январдаги Олий Мажлисга Мурожаатномаси ҳамда бошқа меъёрий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишда ушбу тадқиқот муайян даражада хизмат қилади.

Тадқиқотнинг республика 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.

Тадқиқот 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), Humboldt Universitat (Германия), Boston University, University of New York, University of Pennsylvania, University of California (АҚШ); Indiana University (АҚШ), The University of Sorbonne (Франция), Universitat de Valencia (Испания), University of Zurich (Швейцария), 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), Bakı 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 Крим федерал университети (Крим Республикаси, Россия), шунингдек, Мирзо Улуғбек номидаги

²Диссертация мавзуси бўйича хорижий илмий тадқиқотлар шарҳи google.scholar.com; www.bu.edu; americanenglish.state.gov; micros.uz; www.uv.es; www.zora.uzh.ch; cleee@snu.ac.kr; www.u-tokyo.ac.jp/en; www.ras.ru; adwww.ecu.edu.au; unice.fr/en; www.sophia.ac.jp/eng; www.msu.ru; www.vsu.bu; www.education.ua/universities; www.kaznu.kz; www.navoiy-uni.uz ва бошқа манбалар асосида амалга оширилди.

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 тизими(АҚТ)да лингвистик таъминот, матнга автоматик ишлов бериш технологияси ва воситалари, матннинг автоматик анализ/синтез дастурлари, морфологик анализаторлар тадқиқига оид олиб борилган изланишлардан қуйидаги илмий натижалар олинган: автоматлаштирилган ахборот-қидирув тизимлари лингвистик таъминоти тавсифланган, ахборот қидирув тизимининг асосий ва ёрдамчи воситалари, мултибазавий тил воситалари аниқланган (Cambridge University, Англия; University of Zurich, Швейцария); табиий тилга автоматик ишлов бериш (тил бирликларини анализ/синтез қилиш) хусусиятлари ёритилган (University of New Jersey, University of California, АҚШ; Boston University, New York University, АҚШ; Universitat de Valencia, Испания); табиий тилга компьютерда ишлов беришда трансдюссерлардан фойдаланиш усуллари аниқланган (Courant Institute of Mathematical Sciences, New York, АҚШ; Pereira – San Francisco, USA, АҚШ); матнга автоматик ишлов бериш технологияларининг асбоб-ускуналари ва воситалари ишлаб чиқилган (University of Tokyo, Япония; Australian National University, Австралия; Oxford University Language Centre, Англия); морфологик таҳлилнинг назарий асослари, умумий тавсифи, ишлаш тамойиллари, тилга оид грамматик категория, дастур учун лингвистик модуллар тавсифи ишлаб чиқилган (Kiofo University, Япония; Australian National University, Австралия; Истамбул техника университети, Туркия; Москва давлат университети, Москва давлат лингвистика университети, Санкт-Петербург давлат университети, Россия); матнга морфологик ва синтактик ишлов бериш модели ва дастурлари (Харьков давлат педагогика университети, Украина; Baki Dövlət universiteti, Озарбайжон); туркий тиллар компьютер лингвистикасида табиий тилга ишлов бериш, морфологик анализатор тузиш масалалари тадқиқ қилинган, тил корпуси, морфоанализатор, овозли нутқни синтез қилиш воситалари, электрон таржимонлар ишлаб чиқилган (Татаристон Республика Фанлар академияси амалий тилшунослик ва семиотика институти, Қозон федерал университети, В.Вернадский номидаги Таврия академияси Қрим федерал университети, Россия); ўзбек-инглиз тили машина таржимасининг лингвистик таъминоти, ўзбек тили муаллифлик корпусини, тил корпусининг лингвистик базасини тузиш тамойиллари ва унинг лингвистик таъминоти, ўзбек тили бирликларини графематик таҳлил қилиш муаммолари, ўзбек тили атов бирликларини семантик теглашнинг назарий асослари, ўзбек-инглиз параллел корпуси инструментарийларини яратиш асослари ишлаб чиқилган (Мирзо Улуғбек номидаги ЎзМУ, Алишер Навоий номидаги ТошДЎТАУ, Ал-Хоразмий номидаги ТАТУ, Ўзбекистон).

Дунё компьютер лингвистикасида табиий тилга автоматик ишлов беришнинг қуйидаги устувор йўналишларида илмий тадқиқотлар олиб

борилмоқда: автоматлаштирилган ахборот-қидирув тизимлар лингвистик таъминотини ишлаб чиқиш; табиий тилга автоматик ишлов бериш (тил бирликларини анализ/синтез қилиш) хусусиятларини ёритиш; матнга автоматик ишлов бериш технологияларининг асбоб-ускуна ва воситаларини яратиш; морфологик таҳлилнинг назарий асослари, умумий тавсифи, ишлаш тамойиллари, тилга оид грамматик категория, дастур учун лингвистик модуллар ишлаб чиқиш; туркий тиллар компьютер лингвистикасида табиий тилга ишлов бериш, морфологик анализатор тузиш масалаларини тадқиқ қилиш, тил корпуси, морфоанализатор, овозли нутқни синтез қилиш воситалари, электрон таржимонлар ишлаб чиқиш.

Муаммонинг ўрганилганлик даражаси. Жаҳон тилшунослигида лингвистик таъминотнинг назарий, методологик муаммоларини ўрганиш ўтган асрнинг 60-йилларида бошланди: ахборот-қидирув тилисиз электрон ҳисоблаш машинасининг машиналигича қолиб кетиши тан олинди, “механик” босқичдан “мантикий-лингвистик” босқичга, ахборот-қидирув тизими тараққиёт босқичига ўтилди³. Жаҳон компьютер лингвистикасида табиий тилга ишлов бериш муаммолари А.Кейлер, К.В.Линден, Н.Вард⁴, М.А.Моҳри⁵ томонидан ёритилди. Матнга автоматик ишлов бериш технология ва воситалари В.Ахо Альфред, С.Лам Моника, С.Рави, Д.Ульман Джеффри⁶, морфологик таҳлилнинг назарий асослари С.Ю.Толдова, А.А.Бонч-Осмоловская⁷, Т.Садиков⁸ ишларида тавсифланди. Рус компьютер лингвистикасида О.В.Дереза, Д.А.Каютенко, А.С.Феногенова, М.Ҳакимовнинг морфологик анализатор, унинг умумий тавсифи, ишлаш принциплари, тилга оид грамматик категория, дастур учун лингвистик модуллар тавсифи ҳақидаги ишларини таъкидлаш ўринли⁹. Ахборот-қидирув тизимида лингвистик таъминотнинг аҳамияти А.Б.Антопольский, Ю.М.Арский, Г.Г.Белоногов, Б.А.Кузнецов, В.Н.Белоозеров, Р.Р.Мдивани, В.А.Глинский, А.И.Михайлов, В.А.Мишин, В.В.Морозов¹⁰ ишларида очиб берилган.

³ Соколов А.В. Автоматизация библиографического поиска / А.В.Соколов. – М.: Книга, 1981. – 167 с. – С.91.

⁴ Speech and Language Processing. Daniel S. Jurafsky and James H. Martin. Contributing writers: Andrew Kehler, Keith Vander Linden, Nigel Ward 2000y. Prentice Hall, Englewood Cliffs, New Jersey 07632. Pages: 950.

⁵ Mohri M.A. Finite-state transducers in language and speech processing. Computational Linguistics. – С. 269-312.

⁶ Ахо Альфред В., Лам Моника С, Сети Рави, Ульман Джеффри Д. Компиляторы: принципы, технологии и инструментарий, 2-е изд.: Пер. с англ. – М.: ООО “И.Д. Вильямс”, 2008. – 1184 с.

⁷ Толдова С.Ю., Бонч-Осмоловская А.А. Автоматический морфологический анализ // Фонд знаний «Ломоносов». М., 2011. www.lomonosov-fund.ru/enc/ru/encyclopedia:0127430

⁸ Садыков Т., Кочконбаева Б. Об оптимизации алгоритма морфологического анализа // Шестая Международная конференция по компьютерной обработке тюркских языков «Turklang-2018». (Труды конференции) – Ташкент, 2018. – 320 с.

⁹ Дереза О.В., Каютенко Д.А., Феногенова А.С. Автоматический морфологический анализ для русского языка сравнительный анализ систем // Шестая Международная конференция по компьютерной обработке тюркских языков «Turklang-2018». (Труды конференции) – Ташкент, 2018.; Хакимов М. Логико-лингвистические модели русского языка // Шестая Международная конференция по компьютерной обработке тюркских языков «Turklang-2018». (Труды конференции) – Ташкент, 2018. – 320 с.

¹⁰ Антопольский А.Б. Лингвистическое обеспечение АСНТИ: структура и проблема совместимости // НТИ. Сер.2, 1983. – №7. – С.17-25; Арский Ю.М. Информационные ресурсы для устойчивого развития общества / Ю.М.Арский, А.И.Черный // Междунар. форум по информ. и док. – 2000. – Т. 25, № 1. – С. 3-9.; Белоногов Г.Г. Автоматизированная обработка научно-технической информации. Лингвистические аспекты / Г.Г. Белоногов, Б.А. Кузнецов, А.П. Новоселов // Итоги науки и техники. Сер. Информатика / ВИНТИ. –1984. –Т.8. – 316; Белоозеров В.Н. Ведение и совершенствование Государственного рубрикатора НТИ (ГРНТИ) /

Морфоанализ хусусиятлари Д.Рудольф, морфоанализнинг амаий аҳамияти А.Ермаков¹¹, матнга автоматик морфологик ва синтактик ишлов бериш модели ва дастурлари И.М.Ножов¹² томонидан ўрганилган.

Туркий компьютер лингвистикасида табиий тилнинг компьютер ишлови, жумладан, морфоанализатор тузиш масалалари атрофлича тадқиқ этилди. Бундай ишлар сирасига Г.Эрйигит, Е.Адали¹³, А.В.Дибо, А.В.Шеймович¹⁴, Ж.Сулейманов¹⁵, А.Р.Гатиатуллин, А.М.Баширов¹⁶, П.В.Желтов¹⁷, Н.А.Исраилова, П.С.Бакасова¹⁸, Н.А.Леонтьев¹⁹, В.В.Куканова, А.Ю.Каджиев²⁰ ларнинг морфологик анализатор тузиш масалаларига бағишланган тадқиқотларини киритиш мумкин. Туркий тиллар морфологик таҳлил тизими ҳамда уларнинг моделлари тавсифи, таҳлилига бағишланган тадқиқотлар²¹ кўламининг кенглиги ушбу диссертациянинг назарий

В.Н. Белоозеров, В.И. Федосимов, Г.А.Кривенцова // Информ. продукты, процессы и технологии: НТИ-96: Материалы конф. – М., 1996. – С. 60-69.; Мдивани Р.Р. Лингвистическое обеспечение автоматизированной информационной системы по общественным наукам: состав и структура / Р.Р.Мдивани, В.А.Глинский // Вопр. информ. теории и практики. 1985. – №52. – С. 37-55.; Михайлов А.И. Проблемы лингвистического обеспечения АСНТИ / А.И. Михайлов // Вопр. информ. теории и практики. 1984. – № 51. – С. 5-11.; Мишин В.А. Лингвистическое обеспечение библиотечных автоматизированных систем / В.А. Мишин, Л. Санчес-Мехидо // Сов. библиотековедение. 1981. – № 2. – С. 84-93.; Морозов В.В. Лингвистическое обеспечение автоматизированных систем научно-технической информации / В.В. Морозов, Б.Р. Певзнер. –М.: ИПКИР, 1987. – 76 с.

¹¹ Rodolfe D. Computational Linguistic Text Processing: Lexicon, Grammar, Parsing and Anaphora Resolution. Nova Science Publishers, Inc. – NewYork, 2008. – P. 4-5.; Ермаков А. Морфологический анализатор - основа поисковых систем // <https://www.kv.by/archive/index2004154301.htm>

¹² Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы): диссертация канд. наук. – Москва, 2003. – 190 с.

¹³ Eryigit G., Adali E. An affix stripping morphological analyzer for Turkish \\\ Proceedings of IASTED International Conference ARTIFICIAL INTELLIGENCE AND APPLICATIONS. February 16-18, 2004, Innsbruck, Austria. – P. 299-304.

¹⁴ Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов хакасского и древнетюркского языков / Научное обозрение саяно-алтая рецензируемый научный журнал Номер 2(08), 2014 (Периодичность – 2 раза в год. серия: Филология.Выпуск 2). С. 9-31.

¹⁵ Сулейманов Д.Ш., Гильмуллин Р.А., Гатауллин Р.Р. Морфологический анализатор татарского языка на основе двухуровневой модели морфологии / 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.; Сулейманов Д.Ш., Гатиатуллин А.Р. Модель татарской аффиксальной морфемы и ее реализация // Серия: Интеллект. Язык. Компьютер. – Вып.4. – Казань: Фэн. – 1996. – 113-127.

¹⁶ Гатиатуллин А.Р., Баширов А.М. Морфологический анализатор тюркских словоформ на базе структурно-функциональной модели тюркской морфемы // 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.

¹⁷ Желтов П.В. Разработка морфологического анализатора чувашского 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.

¹⁸ 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.

¹⁹ Леонтьев Н.А. Морфологический анализатор якутского языка // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент, 2018. – 320 с. – С. 276-279.

²⁰ Куканова В.В., Каджиев А.Ю. Алгоритм работы морфологического парсера калмыцкого языка // В сборнике: Писменное наследие и информационные технологии. E'lManuscript-2014 Материалы от V международна научной конференции. Отговорни редактори В.А.Баранов, В.Желязкова, А.М.Лаврентьев. 2014. – С. 116-119.

²¹ Kemal Oflazer. Two-level Description of Turkish Morphology. Literary and Linguistic Computing, – Vol. 9, No 2, – 1994.; Altintas K., Cicekli I. A morphological analyzer for Crimean Tatar // Proceedings of the 10th Turkish

асосларини мустаҳкамлади. Шунингдек, морфоанализатор ишлаб чиқиш назарий ғоялари амалиётга жорий этилиб, Европа, рус, туркий тилларнинг морфоанализаторлари ишлаб чиқилди²². Ўзбек тилшунослигида компьютер лингвистикаси, матнга лексикографик ишлов бериш ва лингвостатистик таҳлил этиш борасида бирмунча тадқиқотлар амалга оширилган. М.Маҳмудов, М.Айимбетов, С.Каримов, Г.Жуманазарова, А.Бабанаров, Д.Ўринбоева, А.Норов, С.Мухаммедов, Н.Абдурахмоноваларнинг тадқиқотларини ана шундай ишлар сифатида қайд этиш мумкин²³.

Ўзбек тили морфологик анализаторини тузиш махсус тадқиқот предмети бўлмаган, аммо айрим изланишларда масаланинг муайян қирралари ҳақида фикр билдирилган. Чунончи, ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш²⁴, ўзбек-инглиз тили машина таржимасининг лингвистик таъминоти²⁵, ўзбек тили муаллифлик корпусини тузиш тамойиллари²⁶, ўзбек тили бирликларини графематик таҳлил қилиш муаммолари²⁷, тил корпуси лингвистик базасини тузиш тамойиллари²⁸ монографик планда ўрганилган. Шунингдек, ўзбек тили атов

Symposium on Artificial Intelligence and Neural Networks (TAINN'2001). – 2001. – P. 180-189.; Çağrı Çöltekin (2014) A Set of Open Source Tools for Turkish Natural Language Processing In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14) Ed. by N. Calzolari et al. 1079-1086.; G. Kessikbayeva and I. Cicekli, "Rule based morphological analyzer of Kazakh language," in Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM. Baltimore, Maryland: ACL, June 2014, pp. 46-54.; Tantug C., Adali E., Oflazer K. Computer Analysis of the Turkmen Language Morphology, 5th International Conference on NLP (FinTAL 2006), Turku, Finland, 186-193, 2006.; Orhun M., Tantug C., Adali E. Rule Based Analysis of the Uyghur Nouns, International Journal of Asian Lang. Proc., 19(1), 33-44, 2009.

²² Морфологический анализатор текста на русском языке mystem [Электронный ресурс]; // Компания Яндекс [сайт] 2003-2013. URL: <http://company.yandex.ru/technologies/mystem/>; <http://tamgasoft.kg/morfo/>; <http://tatmorphan.pythonanywhere.com/>

²³ Маҳмудов 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.; Айымбетов 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 лингвопoeтикаси: 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.

²⁴ Мухаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. – Тошкент, 2006.;

²⁵ 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.

²⁶ Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2018. – 250 б.

²⁷ 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.

²⁸ Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2019. – 140 б.;

бирликларини семантик теглашнинг лингвистик асослари²⁹, ўзбек-инглиз параллел корпуси тузиш муаммолари³⁰, ўзбек тилидаги отларни автоматик таҳлил қилиш³¹, сўз ясалишининг формал моделлари³² ҳақида қатор мақолалар нашр этилган. Ўзбек тилида от ва сифат сўз туркуми морфологик анализини автоматлаштиришнинг дастурий томонлари ўрганилган³³. Бу тадқиқотлар компьютер лингвистикаси муаммоларини тадқиқ этганлиги билан долзарблик касб этган, аммо ўзбек тили автоматик морфологик таҳлил дастури – морфоанализатор ва унинг лингвистик таъминотини яратиш тамойиллари масаласи кун тартибига махсус қўйилмаган. Диссертацияни тайёрлаш жараёнида юқорида санаб ўтилган илмий изланишлар атрофлича ўрганилди, зарур ўринларда уларга муносабат билдирилди 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 адабиёти университетининг АМ-ФЗ-201908172 рақамли “Ўзбек тили таълимий корпусини яратиш” амалий лойиҳаси доирасида бажарилган.

Тадқиқотнинг мақсади ўзбек тили морфологик анализаторининг лингвистик таъминоти назарий асосларини ва моделларини яратишдан иборат.

Тадқиқотнинг вазифалари қуйидагилардан иборат:

лингвистик таъминот воситаларининг умумий тавсифи ҳамда унинг морфологик анализатор ва тил корпуслари яратишдаги аҳамиятини ўрганиш;

тил бирликларининг автоматик морфологик анализ/синтез жараёни воситаларини ўрганиш;

лингвистик таъминотнинг тузилиши, таркиби ва вазифаларини тадқиқ этиш;

морфологик анализатор тузишнинг умумий тамойиллари ҳамда унинг ахборот-қидирув тизимидаги аҳамиятини аниқлаш;

туркий тиллар морфологик анализаторларининг муштарак ва фарқли жиҳатларини тадқиқ этиш;

²⁹ Akhmedova D.B., Mengliev B.R. Semantic Tag Categories in Corpus Linguistics: Experience and Examination International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-3S, October 2019. – P. 208-212.

³⁰ 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.

³¹ Орхун М. Computational analysis of uzbek nouns / Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент, 2018. – 320 с.

³² 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.

³³ Болтаев Т.Б., Ибрагимов С.И. О проекте программной системы морфологического анализа узбекского языка // <http://buxdu.uz/index.php/uz/>; Болтаев Т.Б., Кузьминов Т.В., Поттосин И.В. О структурном конструировании программ и инструментах его поддержки // Среда программирования: методы и инструменты. Новосибирск, 1992, С22-37.; The Structured Constructing as a Discipline of Safe Programming and Instruments Supporting It / Aniskov M.I., Boltaev T.B., Kochetov D.V. at al // Instrumental Congress on Computer Systems and Applied Mathematics CSAM'93. St-Petersburg. July 19-23.

Ўзбек тили морфоанализатори бирликлари, тузилиши, таркиби ва таҳлил бирликларини аниқлаш;

сўзда луғавий шакл ясовчи, синтактик шакл ясовчи, сўз ясовчи морфемалар ҳамда қўшимча кўринишидаги юкламалар қуршовини лингвистик моделлаштириш.

Тадқиқотнинг объекти сифатида ўзбек тили морфологик анализаторининг лингвистик таъминоти танланган.

Тадқиқотнинг предметини лингвистик таъминот бирликларини моделлаштириш ҳамда тизимлаштириш асослари ташкил этади.

Тадқиқотнинг усуллари. Тадқиқот мавзусини ёритишда таснифлаш, тавсифлаш, қиёслаш, статистика, моделлаштириш усулларидан фойдаланилди.

Тадқиқотнинг илмий янгилиги қуйидагилардан иборат:

Ўзбек тилида ахборот қидирув тизимининг лингвистик таъминоти таркиби, вазибаларини аниқлаш орқали орфоқорректор, лемматизатор, морфоанализатор, парсер каби лингвистик анализаторларининг хусусий функционал аспектлари аниқланган;

ахборот-қидирув тизимида морфологик анализаторнинг матни лингвистик аннотациялаш ҳамда машина таржимасидаги табиий тилга автоматик ишлов беришнинг асосий воситаси эканлиги далилланган;

туркий тилларда ишлаб чиқилган морфологик анализаторлар функционал имкониятининг лингвистик таъминот жиҳатидан фарқли, тилга автоматик ишлов беришнинг морфологик таҳлил усуллари жиҳатидан эса муштарак эканлиги аниқланган;

Ўзбек тилидаги имловий ўзгаришлар туфайли сўз шакллариининг умумий қоидаларидан фарқланувчи сўзлар учун морфотактиканинги комбинатор ва позицион лингвистик таъминоти дастури ишлаб чиқилган;

Ўзбек тили бирликларини автоматик морфологик таҳлил қилиш учун сўз ясовчи, луғавий шакл ясовчи, синтактик шакл ясовчи морфемалар ҳамда қўшимча шаклидаги юкламаларнинг сўзга бирикиш позициялари дастурий таъминот алгоритми учун лингвистик моделлаштирилган.

Тадқиқотнинг амалий натижалари қуйидагилардан иборат:

тил бирликлари автоматик морфологик анализ/синтез жараёни воситалари таҳлил қилинган, морфологик анализатор тузишнинг умумий тамойиллари аниқланган, унинг ахборот-қидирув тизими, тил корпуслари, машина таржимаси ҳамда компьютер лексикографиясидаги аҳамияти далилланган;

туркий тиллар морфологик анализаторларининг муштарак ва фарқли жиҳатлари кўрсатилган, ўзбек тили морфоанализатори бирликлари, тузилиши, таркиби тавсифланган, тадқиқот иловасида таҳлил бирликлари рўйхати шакллантирилган;

Ўзбек тили морфологик анализатори ва унинг лингвистик таъминотини ишлаб чиқиш бўйича таклиф берилган;

морфоанализатор учун сўз ясовчи, луғавий шакл ясовчи, синтактик шакл ясовчи морфемалар ва қўшимча кўринишидаги юкламалар 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 морфологик анализаторини тузиш тамойилларини ишлаб чиқишда амалда исботланган манбаларга таянилганлиги билан изоҳланади.

Тадқиқот натижаларининг илмий ва амалий аҳамияти. Тадқиқот натижаларининг илмий аҳамияти ўзбек тили морфологик анализатори яратишнинг назарий асослари, унинг лингвистик таъминотини ишлаб чиқиш доирасидаги янги тадқиқотлар учун илмий-назарий манба бўла олиши, жумладан, тил бирликларининг автоматик морфологик анализ/синтез жараёни воситалари бўйича назарий маълумотлар бериши билан белгиланади.

Тадқиқот натижаларининг амалий аҳамияти амалий тилшунослик, компьютер лингвистикаси, корпус лингвистикаси фанлари ўқитилиш жараёнида дастур, режа тузиш, мавзуларни баён этишда манба вазифасини ўташи, морфологик анализатор тузишнинг умумий тамойиллари, унинг ахборот-қидирув тизими, тил корпуслари, машина таржимаси ҳамда компьютер лексикографиясининг ривожига хизмат қилиши, ўзбек тилидаги матнларни морфологик аннотациялашда қўлланилиши билан изоҳланади.

Тадқиқот натижаларининг жорий қилиниши. Ўзбек тили морфологик анализатори лингвистик таъминоти, морфологик анализатор тузишнинг умумий тамойиллари, морфологик анализаторнинг ахборот-қидирув тизимидаги аҳамияти, ўзбек тили морфоанализатори бирликлари, тузилиши, таркибига доир хулосалардан:

ўзбек тили морфоанализатори бирликлари; тузилиши, таркиби ва таҳлил бирликларини аниқлаш, сўз ясовчи, луғавий шакл ясовчи, синтактик шакл ясовчи ва қўшимча кўринишидаги юкламалар қуршовини лингвистик моделлаштиришга оид натижалардан 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 (Олий ва ўрта махсус таълим вазирлигининг 2020 йил 20 ноябрдаги 89-03-4852-сонли маълумотномаси). Натижада ҳозирги қорақалпоқ тилида функционал сўз ясалиши моделлари мукамаллаштирилган; қорақалпоқ тилидаги шакл ясовчи морфема, қўшимча кўринишидаги юкламалар омонимиясини автоматик фарқлаш тамойиллари ишлаб чиқилган;

ахборот-қидирув тизимида ўзбек тили бирликларини қўллашнинг аҳамияти, ахборот-қидирув тизимининг тил маданиятини шакллантиришдаги ўрни ҳақидаги хулосалардан 2017-2020 йилларда Андижон давлат университетида бажарилган “ОТФ-1-18/ Оммавий лисоний маданиятни

шакллантириш методлари ва методологиясини ишлаб чиқиш” давлат-фундаментал тадқиқотлар дастурлари доирасидаги лойиҳада фойдаланилган. (Олий ва ўрта махсус таълим вазирлигининг 2020 йил 20 ноябрдаги 89-03-4852-сонли маълумотномаси). Натижада оммавий лисоний маданиятни шакллантириш методлари ва методологиясини ишлаб чиқишда ахборот-қидирув тизимида ўзбек тили бирликларини кенг қўллаш тамойиллари ишлаб чиқилган, ахборот-қидирув тизимининг тил маданиятини шакллантиришдаги ўрнини асослашга эришилган;

сўзда бошқа грамматик шакллар билан омонимия ҳосил қилувчи луғавий шакл ясовчи ва синтактик шакл ясовчи морфемалар; контекстда сўз ясовчи морфеманинг омонимик ҳамда полифункционаллик хусусиятини аниқлаш учун сўздаги қуршовни моделлаштириш 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 (Олий ва ўрта махсус таълим вазирлигининг 2020 йил 20 ноябрдаги 89-03-4852-сонли маълумотномаси). Натижада ўзбек ва инглиз тилларида инсон қиёфаси, феъл-атвори, табиат ва миллий тимсоллар тасвири билдирувчи сўзларнинг контекстдаги омонимиясини автоматик аниқлашга эришилган;

ўзбек тили морфологик анализаторининг лингвистик таъминоти, морфологик анализатор тузишнинг умумий тамойиллари, ўзбек тили морфоанализатори бирликлари, тузилиши, таркиби ва таҳлил бирликлари, морфотактика қоидаларини тузиш натижаларидан “Ўзбек тилининг миллий ва муаллифлик корпуси” учун ЭҲМ дастури”ни тузишда фойдаланилган (Ўзбекистон Республикаси Интеллектуал мулк агентлиги томонидан 31.10.2018 санасида берилган № DGU 05735 рақамли гувоҳнома). Натижада ўзбек тили миллий ва муаллифлик корпуси лингвистик таъминоти ишлаб чиқилган, тил корпуси бирликлари, таҳлил бирликлари, теглар тизими сираси тўлдирилган;

сўз ясовчи қўшимчаларнинг қуршовдаги маъносини аниқлаш учун таклиф этилган қўшимчалар аранжировкаси, сўз ясовчи қўшимчаларнинг сўздаги позициясига қараб маъносининг ўзгариши хусусидаги хулосалардан 5220100 – Филология ва тилларни ўқитиш (ўзбек тили) бакалавр таълим йўналиши талабаларига мўлжалланган “Ҳозирги ўзбек тили” дарслигининг “Ясама сўзларнинг лексемалашувчи” деб номланган 13-параграфида ўзбек тили лексикасидаги лисоний ясама сўз ҳамда соф нутқий ясама сўзларнинг ясалиш структурасини тавсифлашда фойдаланилган (Олий ва ўрта махсус таълим вазирлигининг 2020 йил 20 ноябрдаги 89-03-4852-сонли маълумотномаси). Натижада ўзбек тили лексикасидаги лисоний ясама сўз ҳамда соф нутқий ясама сўзларнинг ясалиш структураси тавсифи мукаммаллашган;

ўзбек тили морфологик анализаторининг табиий тилга автоматик ишлов бериш усуллари: стемминг, парсинг, таггинг натижаларидан 5A120106 – Компьютер лингвистикаси магистратура мутахассислигига

мўлжалланган “Korpus lingvistikasi: korpus tuzish va undan foydalanish”. Amaliy mashg`ulot uchun qo`llanma”сида фойдаланилган (Олий ва ўрта махсус таълим вазирлигининг 2020 йил 20 ноябрдаги 89-03-4852-сонли маълумотномаси). Натижада “Корпус лингвистикаси: корпус тузиш ва ундан фойдаланиш” фанини ўқитишда адабиёт ва ахборот манбалари сираси бойитилган;

оммавий саводхонликни ошириш, ёшлар орасида тил маданиятини шакллантириш ҳамда замонавий ахборот тизимида ўзбек тилининг имкониятларидан фойдаланишда, интернетда мавжуд бўлган автоматик таржимон, кирилл-лотин конвентори, компьютер таҳлили, таҳрири, тил корпуси, электрон лугат ва тезаурусларнинг аҳамияти ҳақидаги хулосалардан Бухоро вилояти телерадиокомпанияси «Бухоро» телеканалининг «Ассалом, Бухоро» телекўрсатувида фойдаланилган (Бухоро вилояти телерадиокомпаниясининг 2020 йил 12 ноябрдаги № 1/296-сонли маълумотномаси). Натижада юқори савияли баркамол авлодни вояга етиштириш, ёшлар орасида тил маданиятини шакллантириш ҳамда замонавий ахборот тизимида ўзбек тилининг имкониятларидан фойдаланиш ҳақидаги тарғибот ишлари мазмуни мукаммаллашган.

Тадқиқот натижаларининг апробацияси. Мазкур тадқиқот натижалари 12 та халқаро, 18 та республика анжуманида муҳокамадан ўтказилган ҳамда қуйидаги профилларда ҳаволага эришилган: orcid.org/0000-0002-5429-4708, www.researchgate.net/profile/Shahlo_Khamroeva, <https://scholar.google.com/citations>.

Тадқиқот натижаларининг эълон қилиниши. Диссертация мавзуси бўйича 54 та илмий иш нашр этирилган, жумладан, 2 та монография, Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясининг докторлик диссертациялари асосий илмий натижаларини чоп этиш учун тавсия этилган илмий нашрларда 11 та илмий мақола, улардан 4 таси хорижий журналларда чоп қилинган.

Диссертациянинг ҳажми ва тузилиши. Диссертация кириш, тўрт боб, хулоса, фойдаланилган адабиётлар рўйхати ва иловадан иборат. Ҳажми 264 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 очиб берилган, жорийланиши, апробацияси, нашр этилган ишлар ва диссертация тузилиши бўйича маълумотлар келтирилган.

Диссертациянинг биринчи боби **“Лингвистик таъминотнинг ахборот-қидирув тизимларидаги ўрни”** деб номланган. Бобнинг *“Лингвистик таъминот воситаларининг умумий тавсифи,*

морфоанализатор, тил корпуслари яратишдаги аҳамияти” номли биринчи бўлимида лингвистик таъминот, ахборот-қидирув тизимларида лингвистик таъминотнинг тутган ўрни ёритилган. Жаҳон тилшунослигида лингвистик таъминотнинг назарий, методологик муаммоларини ўрганиш ўтган асрнинг 60-йилларида бошланган. Лингвистик таъминотнинг аҳамияти А.Б.Антопольский, Ю.М.Арский, Г.Г.Белоногов, Б.А.Кузнецов, В.Н.Белоозеров, Р.Р.Мдивани, В.А.Глинский, А.И.Михайлов, В.А.Мишин, В.В.Морозов ишлари³⁴ да очиб берилган.

Лингвистик таъминот – табиий тил матнларини автоматик қайта ишлаш(сўров тили ва қидирув)да қўлланадиган воситалар мажмуи, тил процессори. Лингвистик таъминот – у ёки бу ижтимоий мулоқот доирасида тил воситаларининг адекват ишлашини таъминловчи белгилар мажмуи³⁵, ахборот технологиялари дастурий таъминотида ишлатилувчи тиллар, уларни яратиш усул/воситалари, қўллаш ва фойдаланиш комплекси³⁶. Демак, лингвистик дастурий таъминот ҳақида табиий тилдаги аудиомаълумот, расм ва матнни таҳлил қилиш, қайта ишлаш, сақлашга оид дастур ва маълумотлар мажмуи, деган хулосага келиш мумкин. Ахборот-қидирув тизимининг лингвистик таъминоти таркибига қуйидагилар киради³⁷:

1. Ахборот-қидирув тили (хужжат мазмуни ва қидирув тили): алифбо ва АҚТ микросинтаксиси; мазмун режасини ифодалаш воситаси: луғат, тезаурус, ёрдамчи луғатлар; классификатор; АҚТнинг грамматик воситалари; табиий тил маълумотини ифодалаш воситалари.
2. Турли тилларни мослаштиришни таъминловчи воситалар: мослик жадвали; конвентрлаш жадвали; протоколлар.
3. Амалиёт воситалари: индекслаш методикаси; автоматик индекслаш алгоритми; қидирув воситаси.
4. Мазмуний мутаносиблик критерийлари.
5. Тузилиш тиллари ва маълумотларнинг метатавсифи.
6. Матнга табиий тилда ишлов беришнинг ёрдамчи воситалари.

³⁴ Антопольский А.Б. Лингвистическое обеспечение АСНТИ: структура и проблема совместимости // НТИ. Сер.2, 1983. – №7. – С.17-25.; Арский Ю.М. Информационные ресурсы для устойчивого развития общества / Ю.М.Арский, А.И.Черный // Междунар. форум по информ. и док. – 2000. – Т. 25, № 1. – С. 3-9.; Белоногов Г.Г. Автоматизированная обработка научно-технической информации. Лингвистические аспекты / Г.Г.Белоногов, Б.А.Кузнецов, А.П.Новоселов // Итоги науки и техники. Сер. Информатика / ВИНТИ. – 1984. – Т.8. – С. 316.; Белоозеров В.Н. Ведение и совершенствование Государственного рубрикатора НТИ (ГРНТИ) / В.Н. Белоозеров, В.И. Федосимов, Г.А. Кривенцова // Информ. продукты, процессы и технологии: НТИ-96: Материалы конф. – М., 1996. – С.60-69.; Мдивани Р.Р. Лингвистическое обеспечение автоматизированной информационной системы по общественным наукам: состав и структура / Р.Р.Мдивани, В.А.Глинский // Вопр. информ. теории и практики. 1985. – №52. – С. 37-55.; Михайлов А.И. Проблемы лингвистического обеспечения АСНТИ / А.И.Михайлов // Вопр. информ. теории и практики. 1984. – № 51. – С. 5-11.; Мишин В.А. Лингвистическое обеспечение библиотечных автоматизированных систем / В.А. Мишин, Л.Санчес-Мехидо // Сов. библиотекосведение. 1981. – № 2. – С. 84-93.; Морозов В.В. Лингвистическое обеспечение автоматизированных систем научно-технической информации / В.В.Морозов, Б.Р. Певзнер. – М.: ИПКИР, 1987. – 76 с.

³⁵https://sociolinguistics.academic.ru/915/%D0%9B%D0%B8%D0%BD%D0%B3%D0%B2%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B5_%D0%BE%D0%B1%D0%B5%D1%81%D0%BF%D0%B5%D1%87%D0%B5%D0%BD%D0%B8%D0%B5.

³⁶ <https://works.doklad.ru/view/IP7CrszcUzY.html>

³⁷ https://studopedia.ru/6_11435_ponyatie-i-sostav-lingvisticheskogo-obespecheniya.html

7. Лингвистик таъминотнинг таянч воситалари: электрон луғат тузиш амаллари; электрон луғат билан ишлаш жараёни тавсифи; лингвистик таъминотни ҳужжатлаштириш³⁸.

Бизнингча, лингвистик таъминот – автоматлашган қидирув тизимида қўлланувчи ахборотни автоматик таҳлил қилишни унумли даражага кўтариш мақсадида табиий тилнинг формал қоидаларини яратишга хизмат қилувчи тил воситалари мажмуи. У ахборот объекти тузилиши, унинг ишлаши учун фойдаланиладиган, формаллашган, табиий тил ифодалайдиган тушунча устида бажариладиган амаллар, унинг воситалари таснифи ҳамда тўпламидан ташкил топади.

Бобнинг “Сўз шакллари *нинг* морфологик анализ ва синтези” деб номланган иккинчи бўлимида матн(ахборот)нинг мукаммал лингвистик (лексик, морфологик, синтактик, семантик) таҳлили ҳақида фикр юритилади. Мавжуд энг яхши ахборот-қидирув тизимлари матн ахборотини морфологик таҳлил қилиш вазифасини бажаради, гапнинг синтактик таҳлилинини ҳам амалга оширади³⁹. Морфологик таҳлилнинг вазифаси матндаги ҳар бир сўзнинг туркумга мансублигини автоматик тарзда аниқлаш; сўзнинг қайси лексик-грамматик синфга мансублигини фарқлашдан иборат. Матнни синтактик таҳлил қилиш гапнинг семантик элементларини – гуруҳга мансублик, предикатив асосни ажратишга имкон яратади. Лексик таҳлил – матн ахборотини миллий тил ифодаларини кўрсатувчи алоҳида сўзбоши, гап, сўз шаклида таҳлил қилиш. Тилнинг ўзига хослигини кўрсатувчи гап шакли, лексик ифодани аниқлаб беради. Дастурлаш тиллари формал тил қоидаларига қўшимча киритишни, тўлдирилишни талаб қилади. Табиий тил синтаксиси эса “истиснолар қайдани тасдиқлайди” тамойилида ишлайди⁴⁰. Грамматика – маълум тилнинг гап қурилиш усуллари тавсифи; тилни белгиловчи математик система. Дастурлаш тиллари учун қоидалар системасига асосланилган формал грамматика тавсифи асосида яратилади. Шунингдек, ушбу бўлимда лингвистик ва дастурий таъминотда халқаро стандартлардан фойдаланишнинг аҳамияти ёритилган.

Бобнинг “Лингвистик таъминотнинг тузилиши, таркиби ва вазифалари” номли учинчи бўлимида лингвистик таъминот таркибига кирувчи ахборот базаси тавсифланган. Лингвистик таъминотга асос бўлувчи воситалар сирасига қуйидагиларни киритиш мумкин⁴¹:

1. Лингвистик процессор (*linguistic processors*) – матн ахборотини автоматик қайта ишловчи дастур ҳамда унинг алгоритми; тил воситаларини яратиш ва қўллаш жараёни.

³⁸ Пономарев В.В. Лингвистическое обеспечение и социолингвистическая специфика проблемы автоиндексационной актуализации информационных систем: автореф. диссер. канд. филол. наук. – Москва, 2005.

³⁹ Место и назначение лингвистического обеспечения в информационных системах. Понятие информационной системы // <https://gigabaza.ru/doc/116551-pall.html>

⁴⁰ Ўша манба.

⁴¹ Ўша манба.

2. Лингвистик маълумотлар базаси (*linguistic data bank*) электрон луғат маълумотлар базаси, муҳим ёзув, уларни бошқарувчи дастурий бошқарув воситаларини сақлайди.

Юқоридагилар асосида хулоса қилиш мумкинки, лингвистик таъминот машина таржимаси, компьютер лексикографияси, корпус лингвистикаси, морфологик анализ жараёни учун дастурий таъминотнинг зарурий таркибий қисмидир.

Диссертациянинг II боби **“Морфологик анализатор тузишнинг назарий асослари”** деб номланган. Бобнинг *“Морфологик анализатор тузишнинг умумий тавсифи”* номли биринчи бўлимида морфологик анализатор (МА), морфологик маълумотлар базасини яратиш хусусида сўз юритилади. Морфологик анализатор – луғатдаги сўзшакл ҳамда алоҳида сўзларни қиёслашга мўлжалланган, сўзнинг грамматик хусусиятини ифодаладиган алгоритмлар тўплами⁴². МАнинг асосий вазифаси нотаниш сўзнинг эҳтимолий грамматик маъносини аниқлашдан иборат. Таҳлил қилинаётган сўз луғатдан топилмаса, МА унинг грамматик хусусияти ҳақида эҳтимолий хулосалар бериши керак. Бу вазифанинг бажарилишида бир қанча ахборот, хусусан, таҳлил қилиш қоидалари мажмуини талаб этилади. Матннинг автоматик анализ ва синтези мураккаб амалиётни ташкил қилади: компьютер табиий тилдаги матн таҳлилини белгиланган алгоритм асосида амалга оширади. Автоматик таҳлил давомида компьютер матн ҳақида ўз тилида лексик-морфологик, синтактик, семантик тасаввур ҳосил қилади⁴³.

Матннинг морфологик таҳлилида мавжуд сўзшаклларнинг асосий шакли – лемма ҳамда сўзнинг контекстдаги грамматик маънолари аниқланади. Ўзбек тили агглютинатив тилларга мансублиги боис, кўп ҳолларда, лемма стемга тенг келади. Флектив тилларда бўлганидек, морфоанализ жараёни мураккаб бўлмайди. Кейинги босқич сўзшаклнинг туркумга мансублиги ва морфологик белгиларини топишдан иборат бўлади. Сўзшаклнинг грамматик маъносини билдирувчи қўшимчалар муайян лексеманинг шакл ясовчилари ҳамда унинг парадигма аъзоси сифатида шакллантирувчиси саналади. Таҳлил тизими маълумотлар базасида маълум лексемага бирикиши мумкин бўлган барча қўшимчаларнинг парадигмаси тузилади. Бу жараёнда грамматик, лексик-морфологик омонимлик ҳам кузатилади. Бундай ҳолатда МА сўзшаклга морфологик белгини тўғри бириктиришда гапнинг синтактик ва семантик хусусиятига таянади. МАда таҳлил икки: “ўнгдан чапга” ва “чапдан ўнгга” қоида асосида ишлайди; сўзшаклнинг охириги қисмидаги маъноли қисм – аффиксал морфемани; ундан кейин қўшимчадан ташқари (ўнг томондаги) қисмнинг луғатда мавжудлигини аниқлайди.

Морфологик маълумотлар базасини шакллантириш – МА тузишнинг асосий босқичи. Маълумотлар базаси – фойдаланувчи эҳтиёжига жавоб берадиган кенг кўламли маълумотлар мажмуи: унда маълумот сараланади;

⁴² http://www.solarix.ru/for_developers/docs/morphology_analyzer.shtml

⁴³ Щипицина Л.Ю. Информационные технологии в лингвистике: учеб. пособие / Л.Ю.Щипицина. – М.: ФЛИНТА: Наука, 2013. – 128 с. – С. 43.

жадвал шаклида сақланади. Ҳозирда маълумотлар базаси билан ишлашга мўлжалланган SQL, MySQL, Oracle, Access каби тизимлар мавжуд.

Бобнинг *“Морфологик анализаторнинг ахборот-қидирув тизимидаги аҳамияти”* деб номланган иккинчи бўлимида матнга ишлов бериш ва таҳлил қилиш дастурларининг умумий тавсифи, уларнинг ахборот-қидирув тизимидаги ўрни таҳлил қилинган. Ўрганилган лингвистик таҳлил воситаларининг умумий белгиси сифатида графематик, морфологик таҳлил имкониятлари мавжудлигини таъкидлаш жоиз. Ушбу дастурлар таҳлил даражаси, тилга мослашганлик белгиси билан фарқ қилади, айримида семантик/синтактик таҳлил функциясининг борлиги, айримида ушбу функциянинг етишмаслиги кўринади. Бизнингча, битта таҳлил дастурига бир неча функцияни киритиш дастур афзаллигини оширади, шу сабабли ўзбек тили МАни ишлаб чиқишда таҳлилнинг бир неча босқичини баравар қамраб олишга алоҳида эътибор қаратиш лозим. Шунда яратиладиган ўзбек тили морфоанализатори турли лингвистик таҳлилларни бажарадиган полифункционал анализаторга айланади.

Морфологик таҳлилга мўлжалланган дастурларнинг ўзига хос хусусиятларини ўрганиш матни графематик таҳлил қилиш модули; синтактик луғат; омонимияни аниқлаш ва бартараф этиш модули; дастлабки семантик таҳлил модули; лингвистик қидирув тизими (конкорданс); турли тезаурус ҳамда сўзликдан иборат бўлишини кўрсатди. Масалан, *RHPMorphy*⁴⁴ лемматизация; сўзнинг барча шакллари аниқлаш; сўзга тегишли барча грамматик маълумотларни йиғиш; берилган грамматик хусусиятга мувофиқ сўз шаклини ўзгартириш каби амалларни бажаришга мўлжалланган. Лемматизатор (Мультитран)⁴⁵ ёки рус тилининг морфологик анализатори MyStem⁴⁶ ҳам ўзига хос афзалликларга эга. Ушбу таҳлил дастурларининг умумий хусусияти сифатида морфологик таҳлилни амалга ошириш, лемматизация, сўзшаклни аниқлаш, сўзга тегишли грамматик маълумот бириктириш, бир неча тилни таҳлил қилиш каби афзалликларни таъкидлаш муҳим. Юқорида келтирилган МАларнинг фарқли томони шундаки, айрими статистик, айрими семантик таҳлилни амалга оширади. Славян тиллари МАларини кузатишларимиз МАни ишлаб чиқишда грамматик категория, грамматик шакл, граммема, квазиграммема ва дериватеманинг чегараси аниқ белгилаб олиниши лозим, деган хулосага келишимизга асос бўлди.

Бобнинг *“Туркий тиллар морфологик анализаторларининг муштарак ва фарқли жиҳатлари”* деб номланган учинчи бўлимида туркий тилларда яратилган морфоанализаторлар тавсифи, тадқиқот ва тажрибалар, айрим туркий тилларнинг морфоанализатори шарҳи, ўзбек тилининг морфологик анализи борасидаги изланишлар хусусида фикр юритилади.

Табиий тилга автоматик ишлов бериш масаласи жаҳон тилшунослигида кенг тадқиқ этилган. Туркий тилшуносликда ҳам табиий

⁴⁴ <https://novainfo.ru/article/3394>

⁴⁵ <https://www.multitran.com/c/m.exe?a=4&MessNum=8673&topic=2&l1=4&l2=2>

⁴⁶ <https://yandex.ru/dev/mystem/>

тилнинг компьютер ишлови, морфологик анализатор тузиш масалалари кенг тадқиқ этилган. Бундай ишлар сирасига А.В.Дыбо, А.В.Шеймович⁴⁷, А.Р.Гатиатуллин, А.М.Баширов⁴⁸, П.В.Желтов⁴⁹, Н.А.Исраилова, П.С.Бакасова⁵⁰ ларнинг турли тадқиқотларини киритиш мумкин. Ўзбек тили морфологик анализаторини тузиш махсус тадқиқ этилмаган бўлса-да, масаланинг баъзи томонлари ҳақида фикр билдирилган⁵¹. Туркий тиллар морфоанализаторини тузиш борасидаги кузатишларимиз асосида МАНинг ташкил қилувчилари сифатида қуйидагиларни ажратиш ҳақида хулоса қилдик:

- 1) тил луғати (сўзнинг туркумга мансублик белгиси ҳақидаги изоҳ ҳамда фонологик қоидалар мавжуд луғат);
- 2) сўзшаклининг етарли грамматик тавсифига асосланган (тилнинг автоматик таҳлилига йўналтирилган) компьютер модели;
- 3) маълум қўшимчанинг асосга қўшилиши учун фонетик қоидаларга бўйсунувчи алломорфларини танлаш қоидалари тўплами.

Исталган тилнинг морфоанализаторини тузишда тилнинг ўзига хос хусусиятларига алоҳида эътибор қаратилади. Мутахассислар татар тилини “семантик мураккаб, изчил тузилишли, амалий жиҳатдан автоматик лисоний ҳодиса” сифатида баҳолашади⁵². Татар тили морфоанализаторини ўрганиш асосида унинг АҚТнинг қидирув ва ахборотни қайта ишлаш жараёнида амалий аҳамият касб этиши ҳақидаги хулосага келдик. Чуваш тили МАНинг асосий вазифаси синтактик, семантик таҳлилда чуваш тили сўзшакллариининг морфем тузилиши, морфологик белгиларни аниқлашдан иборат. Чуваш тили МА маълумотлар базаси морфологик белгиларни тавсифловчи типлар рўйхати, луғат, маълумотнома, асос ва қўшимчалар базаси, матн/сўзни

⁴⁷ Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов хакасского и древнетюркского языков / Научное обозрение саяно-алтая. Рецензируемый научный журнал Номер 2(08), 2014 (Периодичность – 2 раза в год. серия: Филология. Выпуск 2). С. 9-31.

⁴⁸ Гатиатуллин А.Р., Баширов А.М. Морфологический анализатор тюркских словоформ на базе структурно-функциональной модели тюркской морфемы // 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.

⁴⁹ Желтов П.В. Разработка морфологического анализатора чувашского 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.

⁵⁰ 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.

⁵¹ Орхун М. Computational analysis of uzbek nouns // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент: Издательско-полиграфический дом «Navoiy universiteti», 2018. – 320 с.; Абжалова М. Матнларга авто-лингвистик ишлов бериш тизимлари // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент: Издательско-полиграфический дом «Navoiy universiteti», 2018. – 320 с.; 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: Издательско-полиграфический дом «Navoiy universiteti», 2018. – 320 с.

⁵² Сулейманов Д.Ш., Гильмуллин Р.А., Гатауллин Р.Р. Морфологический анализатор татарского языка на основе двухуровневой модели морфологии / 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.

сақлаш, сўз таҳлили намунасини ўзида сақловчи ишчи модел структураси маълумотлари, турли сўз туркумларига тегишли сўз таркибининг умумий қонуниятларини қамраб олган⁵³. Айрим ўзига хосликлар билан қирғиз⁵⁴, ёқут⁵⁵ тиллари морфоанализатори ҳам шундай таркибий қисмга эга.

Ўзбек тилининг морфоанализаторини ишлаб чиқишда туркий тилларда яратилган МАлар тажрибаси қўл келади. Ўзбек тили МА монографик планда тадқиқ этилмаган бўлса-да, морфологик автоматик таҳлилнинг айрим масалалари ҳақидаги ишлар⁵⁶ мавжуд. Масалан, МАни конструкциялаш воситалари сирасига лексикон, морфотактика, орфографик қоидалар, ўзак семантикасини шакллантирувчи ахборот структураси бирликлари киритилади⁵⁷. Юқорида кўрсатилган ишларнинг структурасини кузатишларимиз асосида ўзбек тили морфологик анализатори таркибини шакллантириш ҳақида хулоса қилиш мумкин.

Тадқиқотнинг III боби “**Ўзбек тили морфоанализаторини тузишнинг лингвистик асослари**” деб аталади. Биринчи бўлим “*Ўзбек тили морфоанализатори бирликлари, тузилиши ва таркиби*” деб номланади.

Манбаларда автоматик таҳлилнинг стемминг; сўзшаклнинг луғат асосидаги таҳлили; мантикий ёндашув асосидаги таҳлил; жадвал асосида, луғатсиз таҳлил турлари кўрсатилади⁵⁸. Автоматик морфологик таҳлил стемматизация; лемматизация; граммедалаш блокларидан ташкил топади⁵⁹. Стемматизация – қидириляётган сўзнинг асосини топиш жараёни. Стем (сўзшакл асоси) қидириляётган сўзшаклнинг грамматик асосига тенг бўлиши шарт эмас. Лемматизация – сўзшаклни леммага келтириш жараёни. Граммедалаш – сўзшаклга грамматик характеристика ёзиш. Граммема (грамматик характеристика) – сўзшаклнинг маълум морфологик синфга мансублигини кўрсатувчи содда морфологик кўрсаткич⁶⁰. В.А.Плунгян: “граммема – грамматик категория элементларидан бири бўлган грамматик маъно. Бир грамматик категориянинг турли граммедалари бир-биридан фарқланади ҳамда бир пайтда ифодалана олмайди. Грамматик кўрсаткич ҳам граммема. Граммема – морфологик категорияни ташкил этувчи морфологик шакл”⁶¹ деб таърифлайди. Демак, граммема – грамматик маъно, у грамматик категориянинг элементи, грамматик шакл билан ифодаланади.

МАда ишлатиладиган асосий тил бирликлари сифатида қуйидагилар ажратилади: *морф* – элементар сегмент белги; *алломорф* – бир морфеманинг

⁵³ Желтов П.В. Кўрсатилган манба.

⁵⁴ Исраилова Н.А., Бакасова П.С. Кўрсатилган манба.

⁵⁵ Леонтьев Н.А. Кўрсатилган манба.

⁵⁶ Болтаев Т.Б., Ибрагимов С.И. О проекте программной системы морфологического анализа узбекского языка // <http://buxdu.uz/index.php/uz/>

⁵⁷ Ўша манба.

⁵⁸ Марчук Ю. Компьютерная лингвистика. – Москва, 2006. – С. 65.

⁵⁹ Формальные модели и программные инструменты компьютерной обработки татарского языка / Р.Р.Гатауллин, А.Р.Гатиатуллин, О.А.Невзорова, Д.Р.Мухамедшин, Д.Ш.Сулейманов, Б.Э.Хакимов, А.Ф.Хусаинов. – Академия наук РТ, Институт прикладной семиотики АН РТ. – Казань: Изд-во Академии наук Рт, 2019. – 260 с. – С.39-40.

⁶⁰ Кўрсатилган манба.

⁶¹ Плунгян В.А. Классификация морфологических значений // Общая морфология: Введение в проблематику: Учебное пособие. – Изд. 2-е, исправленное. – М.: Едиториал УРСС, 2003. – С. 107. – 384 с.

бир хил фонетик таркибга эга бўлган морфлари йиғиндиси; *сўзшакл* – таркибий қисмлари орасида морфотактика орқали аниқланадиган алоқа мавжуд бўлган алломорфлар кетма-кетлиги⁶²; *морфема* – тилнинг сўзни ташкил этувчи энг кичик маъноли бирлиги; битта вазифани бажарувчи, турли умумий, ўхшаш хусусиятга эга морфлар йиғиндисидан иборат; *бирикиш қоидалари* – МА ишини ташкил этувчи асосий элемент. Демак, ўзбек тили МАни ишга тушириш учун морф, алломорф, сўзшакл, морфема ва бирикиш қоидалари ажратилиши талаб этилади.

МА лингвистик таъминотида асос морфемаларни таснифлаш – муҳим жараён. Анъанавий тилшуносликда сўзларни туркумга ажратиш принциплари маълум. Аммо компьютер лингвистикасида сўзларни туркумга ажратишда бошқача ёндашувни кузатамиз. Сўзлар маълумотлар омборидаги “асослар луғати”да туркумга ажратилган ҳолда берилади, шу сабабли компьютер лингвистикасида сўзларни туркумларга ажратишда семантик, синтактик, морфологик, деривацион, фонологик белгиларга асосланилади⁶³. Омоним сўзлар бир неча маъноси билан бирдан ортиқ туркумга мансуб бўлганлиги сабабли МА тузиш амалиётида асос морфемани фарқлашнинг турли форматлари ишлаб чиқилган:

- 1) асос типини фарқлашда классик сўз туркумига ажратиш;
- 2) асос типини структур-модел морфологик тип асосида фарқлаш.

Биринчи усулдаги тасниф тилшуносликнинг иш жараёнига тегишли бўлса, иккинчи усул дастурлаш жараёнида амал қилади⁶⁴.

Барча асос морфемалар аффиксал морфемаларнинг структур-модел морфологик таснифи асосида N, A, V, D, S каби беш морфологик гуруҳга ажратилади: N – ўнг томонидан исмларга бирика оладиган барча аффиксал морфемаларни бириктирадиган асослар гуруҳи. A – ўнг томонидан сифатга тегишли қўшимчаларни бириктирадиган морфемалар гуруҳи. V – ўнг томонидан феълга бирикадиган қўшимчаларни қабул қилувчи асослар гуруҳи. D – ўнг томонидан N қўшимчалар гуруҳи ҳамда сон шаклини ҳосил қилувчи қўшимчалар бирикадиган гуруҳ. S – ўнг томонидан факат модал аффиксларни бириктира оладиган морфемалар синфи. Санаб ўтилган морфологик гуруҳлар ўз ичида морфемаларни бириктира олишига кўра яна ички гуруҳларга бўлинади⁶⁵.

Туркий тиллар, жумладан, ўзбек тилида сўз контекстдаги позициясига қараб турли маънони англатади⁶⁶. *Нопозицион морфема* – гапдаги ўрнидан қатъи назар, бир семантик гуруҳга мансуб бўладиган морфема. Асос морфема объект маъносини билдирувчи сўзлар гуруҳига кирса, қуршовидан қатъи назар, шу гуруҳга мансуб бўлаверади. *Позицион морфема* гапдаги

⁶² Формальные модели и программные инструменты компьютерной обработки татарского языка / Р.Р.Гатауллин, А.Р.Гатиатуллин, О.А.Неврозова, Д.Р.Мухамедшин, Д.Ш.Сулейманов, Б.Э.Хакимов, А.Ф.Хусаинов. – Академия наук РТ, Институт прикладной семиотики АН РТ. – Казань: Изд-во Академии наук Рт, 2019. – 260 с. – С.39-40.

⁶³ Кўрсатилган манба. – С. 75.

⁶⁴ Кўрсатилган манба. – С. 76.

⁶⁵ Кўрсатилган манба. – С. 76.

⁶⁶ Кўрсатилган манба. – С. 79.

ўрнига қараб турли семантик гуруҳга мансуб бўлади. Бундай морфемаларга барча модда-маъдан отлари кириши мумкин. Масалан: *олтин, ёғоч, темир*. Солиштиринг: *Уста ёғочни арралади // Бозбон гулни ёғоч тувакка ўтқазди*⁶⁷. Позицион асосларга предмет белгисини билдирадиган сўз ҳам, ҳаракат белгисини атайдиган сўз ҳам мансуб бўлади.

МАнинг лингвистик таъминотида асослар луғати (лексикон) бўлиши талаб этилади. Ўзбек тили МА лингвистик таъминоти лексикони учун 85 000 сўзни қамраб олган “Ўзбек тилининг имло луғати”⁶⁸ ёки беш жилдлик “Ўзбек тилининг изоҳли луғати”⁶⁹ асос қилиб олиниши мумкин. Шу луғат сўзликлари асосида лексикон базаси от-лексикон, феъл-лексикон, сифат-лексикон, умумий лексикон (сон, олмош, равиш, боғловчи, кўмакчи, юклама, ундов, модал, тақлид)дан ташкил топиши лозим.

Бобнинг “Ўзбек тили морфоанализаторининг таҳлил бирликлари, сўзшакл ва морфемалар” деб номланган бўлимида ўзбек тили МАда грамматик категория (ГК), грамматик маъно (ГМ), граммема, квазиграммема, дериватемаларнинг фарқланиши тавсифланган.

Грамматик категория (ГК) – муайян маъно умумийлиги асосида бирлашган, ўзаро зидланувчи шакл системаси; у шаклларнинг оддий арифметик йиғиндиси эмас, балки маълум турдаги шаклнинг барқарор муносабати тизмасидан иборат бўлган янги бир бутунлик⁷⁰. Демак, бир хил ГМни ифодаладиган шакллар йиғиндиси *грамматик категория*ни ташкил этади. Морфоанализ жараёнида дастлабки босқич – грамматик шаклнинг қайси ГКга мансублигини аниқлаш, шу сабабли МАда дастлаб тилга хос ГКларнинг фарқланиши талаб этилади.

Ўзбек тилшунослигида эгалик; келишик; сон; нисбат; замон; майл; шахс-сон; даража; бўлишли-бўлишсизлик; ҳаракат тарзи; ўзгалоовчи (феълнинг хосланган шакли) каби 11 номдаги ГК ажратилади⁷¹. Ўзбек тили МА ГКларини белгилашда юқорида саналган 11 та категорияга асосланиш автоматик морфоанализ жараёнининг аниқ натижа беришини таъминлайди.

Ўзбек тилида ГМ қуйидаги усуллар билан ифодаланади:

- 1) асосга ГМ билдирувчи қўшимча қўшиш билан: *китобни*;
- 2) ёрдамчи сўзлар келтириш билан: *Биз келажакка ишонч билан қараймиз*;
- 3) интонация орқали. *Бу китоб* (қайси?) . – Бу – китоб (нима?) каби;
- 4) сўзларни такрорлаш билан.

МА ГМ ифодалашнинг биринчи ва иккинчи усуллари таҳлил қилиш имкониятига эга. Бунинг учун маълумотлар базасида грамматик шакл ва ёрдамчи сўзлар ҳақида маълумот келтириш талаб қилинади. Мавжудлигига кўра ГМ учга бўлинади:

⁶⁷ Кўрсатилган манба. – С. 80.

⁶⁸ Ўзбек тилининг имло луғати. (Тузувчилар: Э.А.Бегматов, А.П.Мадвалиев; Н.Маҳмудов таҳрири остида). – Тошкент: Akademnashr, 2013. – 520 б.

⁶⁹ Ўзбек тилининг изоҳли луғати. 5 жилд. – Тошкент: “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти, 2006-2008. – 680 б.

⁷⁰ Замонавий ўзбек тили. Морфология. (Масъул муҳаррир: Р.Р.Сайфуллаева). – Тошкент: Мумтоз сўз, 2008. – 522 б. – Б. 86.

⁷¹ Замонавий ўзбек тили. Морфология. (Масъул муҳаррир: Р.Р.Сайфуллаева). – Тошкент: Мумтоз сўз, 2008. – 522 б. – Б.186.

1. *Сўзнинг табиатида аввалдан мавжуд ГМ.* Сўз бирор туркумга хос бўлса, маълум ГМни англатади⁷². Демак, ҳар бир сўзга бириктириладиган биринчи граммема унинг табиатида аввалдан мавжуд ГМдан келиб чиқади.

2. *Грамматик восита ёрдамида ҳосил қилинадиган ГМ.* Бундай маъно мустақил сўзда мавжуд бўлиб, грамматик қўшимча ҳамда ёрдамчи сўз воситасида ҳосил қилинади⁷³. Бундан келиб чиқадики, морфоанализда иккинчи граммема сўзга қўшилган *грамматик восита* асосида белгиланади.

3. *Сўз бирикмаси ва гап таркибида вужудга келадиган грамматик маъно.* Мустақил сўз эркин бирикма таркибида бош сўз ёки эргаш сўз, гап таркибида бирор гап бўлаги, ундалма, кириш сўз, ажратилган бўлак, сўз-гап бўлиб келади. Ёрдамчи сўз эса сўз ёки гапни ўзаро боғлайди, қўшимча маъно беради. Бу маъно фақат гап таркибида ҳосил бўлади, гапдан ташқаридаги сўзда бундай ГМ бўлмайди⁷⁴. Бундай ГМ морфоанализ жараёнида таҳлилга тортилмайди, балки бу маъно таҳлилнинг кейинги босқичи – синтактик таҳлилнинг вазифаси. Шу сабабли бундай ГМ граммема сифатида белгиланмайди.

ГКга муносабати, яъни парадигма ҳосил қилиш/қилмаслиги жиҳатидан сўз шакллари икки турга: категориал ҳамда нокатегориал шаклларга бўлинади. Категориал шакл маълум сўз туркумига хос грамматик шакл системасига бирлашувчи, демак, ГК ҳосил қилувчи шаклдир. МАда улар **граммема** сифатида белгиланади. И.А.Мельчук *грамматик маъно* атамасининг алоқа-муносабат шаклига ҳам, нокатегориал шаклларга ҳам, сўз ясовчи шаклларга нисбатан бирдек қўлланилиши тарафдори. Шу сабабли олим грамматик маънони синтактик шакл ясовчилар ва сўз ясовчилардан иборат деб билади: биринчисини **граммема**, иккинчисини **дериватема** деб атади ҳамда бу икки тушунчага хос хусусиятни қуйидагича фарқлайди⁷⁵:

Дериватеманинг тенденциялари	Граммеманинг тенденциялари
1) аниқ; 2) бирикиш имконияти чекланган; 3) ифоданинг ноодатий воситаларига эга; 4) синтактик алоқа доирасига кирмайди; 5) лексик маънога таъсир кўрсатади; 6) кўрсаткичи граммемага нисбатан ўзакка яқин жойлашади; 7) ўзи қўшилган асоснинг туркумини ўзгартиради.	1) мавҳум; 2) бирикиш имконияти чекланмаган; 3) стандарт ифода воситаларига эга; 4) синтактик алоқани бошқаради; 5) лексик маънога таъсир кўрсатмайди; 6) ўзакдан узокда жойлашади; 7) ўзи қўшилган ўзак туркумини ўзгартирмайди.

Нокатегориал шакллар ҳам маълум грамматик маънони англатади, аммо улар ўзаро бир системага бирлашмайди, маълум даражада мустақилликка эга бўлади. Отдаги кичрайтиш, эркалаш, сифатдаги белгининг ортик/камлиги шакли нокатегориал шаклга мансуб. Бундай шакл МАда **квазиграммема** сифатида белгиланади.

⁷² Кўрсатилган манба.

⁷³ Кўрсатилган манба.

⁷⁴ Кўрсатилган манба.

⁷⁵ Мельчук И.А. Курс общей морфологии. Т.1, Москва – Вена: «Языки русской культуры», Венский славистический альманах, Издательская группа «Прогресс», 1997. – С.240 – 287.

И.А.Мельчук квазиграммемани “муҳим оралик ҳодиса” деб таърифлайди. Граммема ва дериватемалар орасидаги чегара мутлақ эмас. Муттасил қўлланувчи, аммо бирор-бир алоқа-муносабат категориясига тааллуқли бўлмаган, муҳим бўлмаган ГМ ҳам мавжуд: доимий қўлланилиши инобатга олинмаса, уларни дериватемаларга киритиш мумкин. Бошқа барча ҳолатда улар граммемага мансуб бўлади. Ҳатто улар орасида локализацияси граммема каби ўзакдан узоқда бўлгани ҳам кузатилади. Бундай грамматик маъно *квазиграммема* деб аталади⁷⁶. Ўзбек тилида грамматик категория ҳосил қилмайдиган барча грамматик шаклларни квазиграммемалар сирасига киритиш мумкин. Отга хос луғавий шакллар: қарашлилиқ шакли *-ники*, ўрин-жой қўшимчаси *-даги*, чегара шакл *-гача*, кичрайтириш-эркалаш шакллари; сифат даражаси кўрсаткичи *-роқ*, ҳаракатнинг кучли/кучсизлигини билдирувчи барча шакллар шу рўйхатга киради.

Туркий тилнинг миллий табиатидан келиб чиққан ҳолда, ўзбек тили грамматик шакллари учга бўлиб ўрганилади⁷⁷. (1) Луғавий шакл ҳосил қилувчилар орасида даража, нисбат категорияси **граммема**, соннинг маъно турларини ҳосил қилувчилар **квазиграммема** сифатида қаралади. (2) Синтактик шакл ҳосил қилувчиларнинг барчаси **граммема** сифатида белгиланади. (3) Луғавий-синтактик шакл ҳосил қилувчи – лексеманинг ҳам луғавий маъносига таъсир қилади, ҳам унинг синтактик вазифасини белгилайди. Сифатдош, равишдош, ҳаракат номи шакли иккиёқлама табиати билан бу гуруҳдан ўрин эгаллайди ва оралик учинчи мақомига эга бўлади⁷⁸. Ушбу шакллар ҳам МАда **граммемалар** сирасига киритилади.

Бобнинг “*Ўзбек тили морфоанализаторида морфотактик қоидалар*” деб номланувчи бўлимида морфотактика, валентлик ва морфотактика қоидаларининг МАдаги аҳамияти тўғрисида баҳс юритилади. МА тузишда яна бир муҳим омил *морфотактик қоидалар* тизимини тузиш саналади. Морфотактика – сўз ясалиш ёки синтактик алоқа-муносабат жараёнида морфемаларнинг валентлик хусусияти, уларнинг морфематик алоқасини ўрганувчи бўлим⁷⁹. Морфотактика морфеманинг бирикувчанлик қобилияти, валентлик хусусияти ҳамда уларнинг морфематик муносабати ўрганилувчи сўз ясалиши бўлимининг таркибий қисмидир⁸⁰. Аффиксал морфемалар ўз қуршовига эга, маълум доирадаги сўз ёки асосларга қўшилади, холос; у ёки бу сўзшаклни ҳосил қилади. Сўз ясалиши ва сўз шакли ҳосил қилишдаги бундай махсус жараён *морфематик муносабат* дейилади⁸¹. Ўзбек тилшунослигида валентлик атрофлича ўрганилган масала⁸² бўлганлиги сабабли бу масалага махсус тўхталмаймиз.

⁷⁶ Кўрсатилган манба.

⁷⁷ Замонавий ўзбек тили. Морфология. (Масъул муҳаррир: Р.Р.Сайфуллаева). – Тошкент: Мумтоз сўз, 2008. – 522 б. – Б. 221-223.

⁷⁸ Кўрсатилган манба.

⁷⁹ <http://www.slovorod.ru/russian-wordform.html>

⁸⁰ http://window.edu.ru/catalog/pdf2txt/516/25516/8411?p_page=10

⁸¹ http://window.edu.ru/catalog/pdf2txt/516/25516/8411?p_page=10

⁸² Абузалова М.К. Субстанциал морфология, валентлик ва синтактик қурилма: Филол. фан. доктори (DSc) диссертацияси. – Самарқанд, 2018. – 224 б.; Мухамедова С.Х. Ўзбек тилида ҳаракат феълларининг семантик ва валентлик хусусиятлари. Докт. дисс. – Тошкент, 2007.; Менглиев Б. Морфологик воситаларнинг

МА морфотактик қоидаларини тузишда сўз ясовчи морфемаларнинг комбинацион моделини ишлаб чиқиш, уларни лингвистик базага киритиш оптимал йўлдир. Зеро, сўз ясовчи морфемалар миқдори салмоқли. Морфотактика қоидаларини тузишда сўз ясовчи аффикслар аранжировкаси ҳам муҳим жараён, сўз ясовчилар локализациясини моделлаштириш – морфотактикадан кейинги босқич. Бизнингча, сўзга ясовчи морфема қўшилганда ўзакда рўй берадиган фонетик ҳодисаларга алоҳида эътибор берилмаса, морфологик таҳлилда иш ҳажми кенгайиб кетади. Морфема асосидаги таҳлил дастурлари⁸³ да морфотактика қоидаларини белгилаш усуллари кузатиш шуни кўрсатдики, дастур базасига тилда мавжуд барча аффиксал морфемалар жойлаштирилади, шундан кейин морфотактика қоидаларини белгилаш функцияси ишга туширилади. Аффиксал морфемалар рўйхатидан маълум бир қўшимча танланади. Шу қўшимчадан кейин жойлаша оладиган морфема белгиланади, сақлаш тугмаси босилади. Шундай тартибда қўшимча морфеманинг ҳар бирига ўнг томонидан ёндош бўла оладиган морфемалар бириктирилади. Кузатишларимиз асосида ушбу усул морфотактика қоидаларини тузишда энг қулай йўл эканлигини аниқладик.

Компьютер лингвистикасида сўз *танловчи* ҳамда сўз *танламас* морфемалар фарқланади⁸⁴. Айрим қўшимчалар бир морфологик гуруҳга мансуб барча ўзакка қўшила олади. Масалан, -га қўшимчаси N морфологик гуруҳидаги ўзакларга қўшилади. Бундай қўшимча сўз *танламас морфема* дейилади. Бир морфологик гуруҳга мансуб ўзакларнинг барчасига қўшила олмайдиган қўшимчага ўзлик ва мажҳул нисбат қўшимчалари (-ин, -н, -ил, -л) киради. Улар потенциал жиҳатдан феъл-асос морфемага қўшилади, аммо амалда маълум сўзлар бу қўшимчани қабул қилолмайди. Масалан, *безамоқ* ўзаги иккала нисбат қўшимчасини ҳам қабул қилади: *безанмоқ*, *безалмоқ*. Бу қўшимчаларнинг ўзакка қўшилиш/қўшилмаслигида бирор қонуният мавжуд эмас. Нисбат қўшимчаларининг асосга қўшилиш/қўшилмаслик имкониятини аниқлаш эса морфоанализ жараёнини соддалаштиради, синтезни тезлаштиради. Демак, ўзбек тилида айрим сўзшаклларга қўшиладиган граммемалар борки, уларни сўз *танловчи қўшимчалар* сирасида алоҳида тартиблаштириш лозим, акс ҳолда, морфологик таҳлил жараёни тўғри натижа бермайди.

Диссертациянинг IV боби “**Ўзбек тили морфоанализатори учун сўзларда қўшимчалар қуршовини лингвистик моделлаштириш масалалари**” деб номланган. Бобнинг “*Сўзда лугавий шакл ясовчи, синтактик шакл ясовчи ва қўшимча кўринишидаги юкламалар қуршовини лингвистик моделлаштириш*” деб аталувчи бўлимида ЛШЯ, СШЯ ва қўшимча кўринишидаги юкламалар қуршовини моделлаштириш муаммолари

маъновий хусусиятлари ва семантик имкониятлари. Филол. фан. номз. дисс. автореф. – Тошкент, 1996; Шаҳобиддинова Ш. Грамматик маъно талқини хусусида. Номз. дисс. – Самарқанд, 1993.

⁸³ <http://tatmorph.pythonyanywhere.com/> (25.11.2020)

⁸⁴ Формальные модели и программные инструменты компьютерной обработки татарского языка. / Р.Р.Гатауллин, А.Р.Гатиатуллин, О.А.Неврозова, Д.Р.Мухамедшин, Д.Ш.Сулейманов, Б.Э.Хакимов, А.Ф.Хусаинов. – Академия наук РТ, Институт прикладной семиотики АН РТ. – Казань: Изд-во Академии наук РТ, 2019. – 260 с. – С.81-83.

таҳлил этилади. Грамматик воситанинг сўз таркибидаги тартиби маълум қонуниятга эга⁸⁵. Шу сабабли ЛШЯларнинг сўз туркумларидаги қуршови алоҳида моделлаштиришни талаб этади.

Феълга хос ЛШЯ бўлган *-(u)ш* шакли ҳаракат номи ва биргалик нисбат вазифасини бажаради. Бундай грамматик омонимиянинг фарқланиши учун маълумотлар базасига қуйидаги позицион моделлар киритилиши керак:

- 1) $V+f_1^v$ ёки $V+f_1^v+f_2^v$ позициясида *-(u)ш* шакли ҳаракат номи вазифасини бажаради: *чиқ+ар+иш*; *том+из+иш*, *кий+гиз+иш*, *кий+дир+иш*;
- 2) $V+f_1^v+p_1^v$ ёки $V+f_1^v+f_2^{v+}p_1^v$ позициясида *-(u)ш* шакли биргалик нисбат вазифасини бажаради: *кел+иш+ди*, *суриштир+иш+ди*, *ўтир+иш+ди*.

Феъл + биргалик н. + биргалик н.: *сўзла+ш+иш*. Бу позицияда *-ш* биргалик нисбат шакли ҳисобланиши лозим.

3) *жойлашди*, *керишди*, *қоришди* сўзшаклларида *-(u)ш* ўзлик нисбат маъносини беради, бу феълларнинг ўтимли/ўтимсизлик муаммолари билан боғланиб кетади (албатта, бу муаммо назарий аспектда ўз ечимини кутмоқда)⁸⁶; шунинг учун бу ҳолатни моделлаштириш асосида эмас, балки истиснолар қонидаси асосида аниқлашга тўғри келади.

Феълнинг равишдош шакли позициясига қараб *-(u)б* ўтган замон шакли билан омонимия ҳосил қилади. Позицияларни таҳлил қиламиз:

- 1) *ўқиб*, *бориб* шаклида равишдош кўшимчаси; *ўқиб кўрмоқ*, *гуллаб қўймоқ* шаклида кўмакчи феълли сўз кўшилмасида етакчи ҳамда кўмакчи феълни боғлашга, *олиб бормоқ*, *кириб чиқмоқ* кўшма феълларида етакчи ва ёрдамчи феълни боғлашга хизмат қилган равишдош шакли бўлади. Демак, $V+f_1^v(и)б$, яъни {феъл-асос + (и)б + 0} бўлса, *-(u)б* = феълнинг равишдош шакли; агар $V_1 + f_1^v(и)б + V_2 + f_1^v + f_1^v/p_1^v$, яъни {феъл-асос + (и)б + феъл-асос + ЛШЯ/СШЯ} бўлса, *-(u)б* = етакчи ҳамда кўмакчи феълни боғловчи равишдош шакли;
- 2) *ўтирибман*, *борибсан*, *узоқлашибмиз* сўзларида ўтган замон аффикси вазифасини бажаради. Ушбу позициясини қуйидагича моделлаштирамиз: агар $V+f_1^v(и)б+p_1^v$, яъни {феъл-асос+(и)б+СШЯ} бўлса, *-(u)б* = ўтган замон шакли.

Шолча, *боғча* (*болалар боғчаси*), *кўрпача* сўзларида янги лексик маъно вужудга келганлиги учун *-ча* ЛШЯ ҳисобланмайди, сўзлар маъноли қисмга ажралмайди⁸⁷. Бундай сўзлар истиснолар луғатига киритилади; *-ча* морфем таҳлилга тортилмайди. Ўнгдан чапга таҳлилда *-ча* аффикс сифатида топилиши мумкин, аммо ўзак қисми текширилганда, истиснолар луғатида бу сўзлар асос сифатида топилади; таҳлил ушбу сўзларнинг битта леммага тенг эканлигини аниқлаш билан якунланади.

-лар кўшимчаси ҳурмат маъносини ифодалаганда эгалик шаклидан кейин, соф кўпликни билдирган эгалик шаклидан олдин жойлашади. Қиёсланг: *опамлар // опаларим*. Шунингдек, *опаларим* сўзшаклида *-лар* шахсга нисбатан умумийликни, *опамлар* сўзшаклида конкретликни

⁸⁵ Mengliyev B., Xoliyrov O., Abdurahmonova N. O'zbek tilidan universal qo'llanma. (Qayta ishlangan 3-nashri). – Toshkent: Akademnashr, 2014. – 389 b. – B.134.

⁸⁶ Кўрсатилган манба.

⁸⁷ Кўрсатилган манба. – Б. 175.

билдиради. Морфоанализда ушбу ҳолатнинг фарқланиши қуйидаги модел билан таъминланади. $f = -\text{лар}$ бўлганда: 1) агар $\{R+f+r\}$ бўлса, $f = \text{кўплик}$; 2) агар $\{R+r+f\}$ бўлса, $f = \text{хурмат маъноси}$.

Ҳозирги ўзбек адабий тилида мавжуд шахс-сон кўрсаткичининг қуйидаги тартиби кузатилади⁸⁸:

- 1) феъл-асос + $\{-\text{ган} / -(\text{и})\text{б} / -\text{яп} / -\text{а} / -\text{й} / -\text{моқчи}\} + \{-\text{ман} / -\text{сан} / -\text{о} / -\text{миз} / -\text{сиз} / -(\text{ди})\text{лар}\}$.
- 2) Феъл-асос + $\{-\text{ди} / -\text{са}\} + \{-\text{м} / -\text{нг} / -\text{о} / -\text{к} / -\text{нгиз} / -(\text{ди})\text{лар}\}$.
- 3) Феъл-асос + $\{-\text{ай} / -\text{о} / -\text{син} / -\text{айлик} / -\text{гин} / -(\text{и})\text{нгиз} / -(\text{син})\text{лар} / +\text{-ишсин}\}$.

Ушбу қўшимчалар қуршовини морфоанализ учун қуйидагича моделлаштирамиз:

- 1) $V+f_1^v+r_1^f$. Бунда: $f_1^v = \{-\text{ган} / -(\text{и})\text{б} / -\text{яп} / -\text{а} / -\text{й} / -\text{моқчи}\}$; $r_1^f = \{-\text{ман} / -\text{сан} / -\text{о} / -\text{миз} / -\text{сиз} / -(\text{ди})\text{лар}\}$;
- 2) $V+f_2^v+r_2^f$. Бунда: $f_2^v = \{-\text{ди} / -\text{са}\}$; $r_2^f = \{-\text{м} / -\text{нг} / -\text{о} / -\text{к} / -\text{нгиз} / -(\text{ди})\text{лар}\}$;
- 3) $V+r_3^f$. Бунда: $r_3^f = \{-\text{ай} / -\text{о} / -\text{син} / -\text{айлик} / -\text{гин} / -(\text{и})\text{нгиз} / -(\text{син})\text{лар} / +\text{-ишсин}\}$.

Кўринадик, исм ва феълларга хос синтактик шакл ясовчи морфемалар грамматик омонимияни ҳосил қилади. Синтактик шакл ясовчи доим сўзнинг охирида келганлиги сабабли МАда ўнгдан чапга қидирувда уларнинг айрилиқ шаклдош бўлиб қолади, яъни анализатор маълумотлар омборидаги маълумотга таяниб, *укангиз* ва *келдингиз*, *бордим* ва *укам*, *келсанг* ва *уйқунг* сўзларида *-нгиз* шаклини синтактик шакл ясовчи деб топади. Шу ерда савол туғилади: ушбу синтактик шакл ясовчи отга хосми ёки феълга хосми? МА жавоб топиши учун иккинчи алгоритмни ишга солади, бунинг учун модел талаб қилинади. *-нгиз* шакли синтактик шакл ясовчиларнинг қайси кўриниши эканлиги қуйидаги модел ёрдамида аниқланади:

- 1) агар $N / A / \text{Num} / \text{Pr} + r_1^n$, яъни $\{\text{исм-асос} + -(\text{и})\text{м} / -\text{инг} / -(\text{и})\text{миз}\}$ бўлса, $r^n = \text{эғалик шакли}$;
- 2) агар $V + r^f$, яъни $\{\text{феъл-асос} + -(\text{и})\text{м} / -\text{инг} / -(\text{и})\text{миз}\}$ бўлса, $r^f = \text{шахс-сон кўрсаткичи}$.

Бобнинг иккинчи бўлими “Сўз ясовчи морфемаларнинг сўздаги қуршовини моделлаштириш” деб аталади. Туркий тилдаги сўзларнинг структураси мураккаб бўлишига қарамай, ўзак ва қўшимчадан иборат. Сўз ясалиши, сўз асоси морфологиясининг айрим масалаларини Н.Д.Арутюнов⁸⁹, сўзнинг морфологик структурасини Н.А.Баскаков⁹⁰, Э.В.Севортян⁹¹, турк тилида сўзнинг морфологик структурасини М.С.Джикия⁹², туркий

⁸⁸ Замонавий ўзбек тили. Морфология. (Масъул муҳаррир: Р.Р.Сайфуллаева). – Тошкент: Мумтоз сўз, 2008. – 522 б. – Б. 420-426.

⁸⁹ Арутюнов Н.Д. Некоторые вопросы образования и морфологии основ слов. – Филологические науки, 1958. – № 1.

⁹⁰ Баскаков Н.А. Морфологическая структура слова и части речи в тюркских языках. – Советское востоковедение, 1957. – № 1.; Баскаков Н.А. Порядок и иерархия аффиксов в основе слова в тюркских языках. Баскаков Н.А. В сб.: Морфологическая типология и проблема классификации языков. М.-Л., 1965.

⁹¹ Севортян Э.В. Морфологическое строение слова в связи с другими его характеристиками. – В. Кн.: Тюркологический сборник. – М., 1972.

⁹² Джикия М.С. Морфологическая структура слова в турецком языке: автореф. дисс. ... канд. филол. наук. – Тбилиси, 1975.

агглютинация табиатини А.Н.Кононов⁹³ ўрганган. МАнинг ясама сўз, ясовчи қисм ва ясалиш асосини аниқлай олиши учун унинг лингвистик таъминотида ясама сўз структураси ҳақидаги мукамал маълумот талаб қилинади. Ясама сўз структурасида маъноси ва вазифасига кўра ўзаро фарқланадиган икки хил компонент: сўз ясалиш асоси ҳамда сўз ясовчи шакл мавжуд бўлади. Ясама сўзни ҳосил қилиш асос бўладиган қисм сўз ясалиш асоси ҳисобланади: *туртки, ишчан, қорай* сўзларида *турт, иш, қора* сўз ясалиш асосидир⁹⁴.

Анализаторнинг асослар луғатида ўзбек тилидаги барча ўзаклар туркумга тегишлилик ҳақидаги маълумот билан сақланади, яъни барча ўзаклар леммаланган, қайси сўз туркумига тегишлилиги ҳақидаги ахборот (тег) бириктирилган бўлади.

Морфемани ўнгдан чапга қараб аниқлашда МА уч нарса: сўзнинг асоси, ясалиш асоси ва сўз ясовчи аффиксни аниқлаш қуйидаги алгоритмда амалга ошади:

- 1) ўзак асослар луғатида мавжуд бўлади: асос топилади;
- 2) маълумотлар базасида ясовчи қўшимчалар ҳақидаги маълумотга асосланиб, ўнгдан чапга таҳлил қилинади: сўз ясовчи қўшимча аниқланади;
- 3) ясовчи асосдан олдинги қисм ясалиш асоси деган хулоса келиб чиқади: ясалиш асоси топилади⁹⁵.

Сўз ясалишининг асосий инвариант моделини $R+d$ деб белгилаймиз. Бунда R – ҳар қандай асос; d – дереватема, яъни сўз ясовчи морфема. Асосий $R+d$ инвариант модели R_v+d ; R_n+d ; R_a+d ; $R_{num}+d$; $R_{pr}+d$; $R_{adv}+d$; $R_{sim}+d$ каби субмоделларда воқеалашади.

Терим, гавдали, гулла, кийимлик, бирлашма, теримчи, теримчилик; асаларичи, белбогли; беш йиллик, темирйўлчи, ўзибўларчилик каби ясама сўзлар МА таҳлилида ўнгдан чапга *-им, -ли, -ла, -лик, -ма, -чи, -чилик, -арчилик* каби сўз ясовчи морфемаларни аниқлаши керак. Бир бўғиндан иборат қўшимчаларни аниқлашда муаммо бўлмайди, аммо *-чилик, -арчилик* шакллариининг ўнг томондан ўқирилишида дастлабки *-лик* шакли ажратиб олинishi мумкин. Шунинг учун маълумотлар омборига қоида ва модел киритилади. Агар {ўзак + лик} бўлса, *-лик* сўз ясовчи бўла олади. Моделлари:

- 1) агар $\{R + d_1(лик)\}$ бўлса, $лик = d$;
- 2) агар $\{R + d_1 + d_2(лик)\}$ бўлса, $d_1 + лик = d$.

Отдан от ясалиши инвариант R_n+d^n , отдан сифат ясалиши инвариант R_n+d^a модели, оддан равиш ясалиши инвариант R_n+d^{adv} модели, оддан феъл ясалиши инвариант R_n+d^v , сифат-асосдан ясалиш инвариант R_a+d модели, сон-асосдан ясалиш инвариант $R_{num}+d$, олмош-асосдан ясалиш инвариант $R_{pr}+d$, равиш-асосдан ясалиш инвариант $R_{adv}+d$, тақлид-асосдан ясалиш инвариант $R_{sim}+d$, феъл-асосдан ясалиш инвариант R_v+d модели билан белгиланади. Айтиш жоизки, МАнинг сўз ясалиш таҳлилида фақат морфологик шакллариининг аниқланиши, яъни ясовчи қисмни аниқлаш билан

⁹³ Кононов А.Н. О природе тюркской агглютинации. – Вопросы языкознания, 1976. – № 4.

⁹⁴ Кўрсатилган манба.

⁹⁵ Степанова М.Д. Вопросы моделирования в словообразовнии и условия реализации моделей. – Вопросы языкознания, 1975. – № 4.

сўз ясалишининг тўлиқ таҳлилига эришиб бўлмайди. Сифатли таҳлилга эришиш учун сўз ясалиш типлари ҳам таҳлил қилиниши лозим. Шундан келиб чиқиб, тадқиқотда инвариант моделларнинг жами 39 та субмоделли ишлаб чиқилди, чунки МАда грамматик таҳлил субмоделлар асосида олиб борилади.

$R+d_1+d_2$ ва бошқа мураккаб субмоделдаги сўз ясовчи морфемалар қурилови. Ўзбек тилида бир асосга бирдан ортиқ сўз ясовчи морфема қўшилиши мумкин: бунда сўз ясовчи морфемаларнинг ўзаро тартиби мавжуд. Асосий инвариант $R+d_1+d_2$ қуйидаги субмоделлар орқали воқеланади: $R_n+d_1+d_2$; $R_a+d_1+d_2$; $R_{num}+d_1+d_2$; $R_{pr}+d_1+d_2$; $R_{adv}+d_1+d_2$; $R_{sim}+d_1+d_2$; $R_v+d_1+d_2$. Бунда феъл, от, сифат, сон, олмош, равиш, тақлид асосга кетма-кет икки сўз ясовчи морфема қўшиш билан сўз ясалади. Моделлар МАнинг шаклини таниши, асос туркумини аниқлаши, асосдан кейин қўшилган морфемалар тартиби ҳамда бу тартибнинг маънога таъсирини таҳлил қилиши нуқтаи назаридан ишлаб чиқилди.

Бобнинг “Сўз ясовчи қўшимчалар қуриловини моделлаштиришнинг айрим муаммолари хусусида” деб номланувчи бўлимида сўз ясалиш моделларидаги истиснолар, омоним, полисемантик сўз ясовчи морфемаларни лингвистик моделлаштириш масалалари таҳлилга тортилади.

Ўзбек тилида сўз ясалишининг баҳсли масалалари⁹⁶ махсус таҳлил этилган бўлиб, МА учун моделларни шакллантиришда ушбу ишлардан ўринли фойдаланилди. А.Ҳожиёв ўринли таъкидлаганидек, сўз ясалиши тадқиқида ясовчининг жойлашуви ҳодисани ўрганишда аҳамиятсиз бўлиши мумкин, аммо морфоанализ учун ясовчи қисмнинг позицияси аҳамиятли. Сўз ясовчи қўшимча асосдан кейин, яъни ўнг томонда жойлашган бўлса, анализатор ўнгдан чапга таҳлил қилишда сўз ясовчи аффиксни аниқлайди. Масалан: *супурги*, *кийим*, *сузма* каби ясама сўзларни таҳлил қилар экан, ясалиш асоси ҳамда ясовчи қисмни маълумотлар омборидаги ахборотга асосланиб *супур+ги*, *кий+им*, *суз+ма* тарзида аниқлайди, натижани кўрсатади. Аммо анализатор маълумотлар омборида {сўз ясовчи + ясалиш асоси} қолипи бўлмаганда ясовчи асосдан олдин жойлашган сўз ясовчи (олдқўшимча)лар таҳлилини амалга оширишда натижа беролмайди, шунинг учун МАда бундай ясалиш типи алоҳида формула билан белгиланиши лозим. Демак, аффиксоид ва аффикслар рўйхати маълумотлар омборидан жой олса, МА таҳлилни моделлар асосида олиб бора олади.

Қатор қўшимчалар омонимияси маълум сўз/сўз шаклдаги функциясини модел орқали аниқлаш мумкин: *қорай* ва *ўқий* сўзларидаги *-й* омоним қўшимча, у феъл ясовчи ва феълга хос синтактик шакл ясовчи бўла олади. Маълумки, ўзбек тилида феъл ясовчи қўшимча феъл бўлмаган асосдан феъл ясайди. Шу қоидага асосланиб ушбу шакл позициясини моделлаштирамиз: агар {феъл бўлмаган асос + *-й*} бўлса, *й* = феъл ясовчи; агар {феъл-асос + *-й*} бўлса, *й* = луғавий шакл ясовчи. Демак, модел асосида филтрлаш йўли

⁹⁶ А.Ходжиев, А.Нурманов, Н.Махмудов. Спорные вопросы в системе словообразования тюркских языков. // Советская тюркология, 1989. – №2.

топилди. *-кин* морфемаси *кескин* сўзида сифат ясовчи, *тўккин* сўзида феълга хос ЛШЯ вазифасини бажаради; икки ҳолатда ҳам феъл асосга қўшилган. Асосга нисбатан позициясини модел орқали аниқлаб бўлмайди.

Сўздаги морфемалар қуршови ҳамда тартибини моделлаштиришда қуйидаги хулосаларга асосланиш мумкин:

- 1) сўзда морфемаларнинг маълум тартиби мавжуд; уларда одатдаги тартиб қуйидагича: асос + сўз ясовчи морфема + ЛШЯ морфема + СШЯ морфема;
- 2) сўз ясовчи морфемалар қуршови ўзбек тилининг агглютинативлик хусусияти билан боғлиқ ҳамда уларнинг кетма-кет қўшилиш ҳолатлари кўп, ички ва ташқи сўз ясалиш ҳам фарқланади;
- 3) ўзбек тилида таркибида 4 та сўз ясовчи учрайдиган сўзлар кам бўлса ҳам учрайди, бу лингвистик иқтисод қилиш принципига амал қилмасликдан келиб чиқади;
- 4) барча сўз ясовчилар асосдан кейин, ЛШЯдан олдин жойлашади;
- 5) агар сўз таркибида иккита сўз ясовчи мавжуд бўлса, аввал аниқ от ясовчи, кейин мавҳум от ясовчи қўшилади;
- 6) ўзлашма сўз ясовчилар ўзбек тилидаги сўз ясовчидан кейин қўшилади;
- 7) кучсиз семантик характерли сўз ясовчи сўзда сўнг позицияни эгаллайди;
- 8) каммаҳсул ясовчи олдин, сермаҳсул сўз ясовчи кейин жойлашади;
- 9) сўз таркибида кичрайтириш-эркалаш қўшимчаси бўлганда, аввал кичрайтириш, кейин эркалаш шакли қўшилади;
- 10) иккита функционал бир хил ЛШЯ кетма-кет қўшилади, яъни плеонастик характер касб этади;
- 11) феъл луғавий шакл ясовчиларининг тартиби қуйидагича бўлади: {нисбат шакли + бўлишсизлик шакли + феълнинг функционал шакллари + замон шакли};
- 12) СШЯлар ҳақиқий сўз шаклини ҳосил қилувчи саналади, чунки улар сўзнинг синтактик вазифасини шакллантиради;
- 13) ўзбек тилида қўшимча шаклидаги юкламалар доим сўзнинг охирига қўшилади (уларнинг ҳам кетма-кет қўшилиш ҳолатлари учрайди);
- 14) сўз таркибида одатдаги тартибнинг бузилиши маълум лингвистик факторлар билан боғлиқ бўлади.

ХУЛОСА

1. Лингвистик таъминот ахборот технологиялари дастурий таъминотида ишлатилувчи тиллар, уларни яратиш усул, восита, қўллаш ҳамда фойдаланиш комплекси. Табиий тилда матнларни автоматик қайта ишлашда қўлланиладиган воситалар мажмуи, тил процессори бўлса, ахборот тизимларининг лингвистик таъминоти ахборот объекти тузилиши, унинг ишлаши учун фойдаланиладиган, формаллашган, табиий тил ифодаладиган тушунча устида бажариладиган амаллар, унинг воситалари тўплами ҳисобланади.

2. Лингвистик дастурий таъминот деганда табиий тилдаги аудиомаълумот, расм, матнни таҳлил қилиш, қайта ишлаш, сақлашга оид дастур ва маълумотлар мажмуи назарда тутилади. Унинг электрон луғат, орфоқорректор, кидирув тизими, машина таржимаси системаси, таржимон хотирасини бошқариш дастури, нутқ анализ/синтези тизими, овозли таржима тизими каби турлари ажратиладики, бунда ахборот-кидирув тизимининг лингвистик таъминоти таркибига алифбо, микросинтаксис, луғат, тезаурус, классификатор; турли тилларни мослаштиришни таъминловчи воситалар; амалиёт воситалари киради.

3. Морфологик таҳлилнинг вазифаси матндаги сўзнинг туркумга мансублик белгисини автоматик тарзда белгилаш, лексик-грамматик синфга мансублигини аниқлашдан иборат. Матн ахборотини синтактик таҳлил қилиш гапнинг семантик элементи(гуруҳга мансублик, предикатив асос)ни ажратишга имкон яратади. Лексик таҳлил тилнинг ўзига хослигини кўрсатувчи гап шакллари, лексик ифода (ўзлашма, жаргон сўз)ни аниқлайди. Шунингдек, автоматик таҳлил давомида компьютер матн ҳақида ўз тилида лексик-морфологик, синтактик, семантик тасаввур ҳосил қилади. Ахборот базаси, таҳлил дастурларининг амалий аҳамияти АҚТ ҳамда машина таржимаси сифатини яхшилаш, тил корпуси бирликларини лингвистик разметкалаш, шунингдек, мустақил анализатор сифатида ишлаши билан белгиланади. Зеро, морфологик анализатор луғатдаги сўзшакл, алоҳида сўзларни қиёслашга мўлжалланган, сўзнинг грамматик хусусиятини ифодаладиган воситадир.

4. Морфоанализаторнинг лингвистик таъминоти маълумотлар базаси кўринишида сақланади: база асослар луғати ва қоидалар луғатидан иборат. Маълумотлар базаси қўламли маълумотлар йиғиндиси бўлиб, унда маълумот сараланади, база тизимида жадвал шаклида сақланади. Асослар луғати тил лексикасидан ташкил топади, грамматик луғатда тилдаги грамматик шакллар рўйхати тўпланади; аффикслар базаси; таснифлаш қоидалари; алломорфлар мослиги қоидалари маълумотлар базасидан жой олади. Морфоанализатор лингвистик таъминоти ўзбек тилидаги флексия ҳолатлари (ўзак ва қўшимча чегарасидаги фонетик ўзгаришнинг акс этиши); одатий синтактик шакл

ясовчилар билан бирга супплетив, ўзгармас лексемалар ҳақидаги маълумотни камраб олиши; омоним лексемалар базасини ўзида мужассамлаштиради.

5. Морфоанализатор тузиш учун тилнинг грамматик категория, грамматик шакллари тавсифланади, граммема, квазиграммема ва дериватеманинг чегараси аниқ белгиланади. МА тизимига сўз туркуми ва бошқа морфологик белгиларини аниқлаш; ўзак/асос ва қўшимчани фарқлаш; матнни тушуниш; истисно сўзларни таҳлил қилиш; натижани файл кўринишида тақдим қилиш талаблари қўйилади. Билъакс, морфоанализаторнинг асосий бирликлари морф, алломорф, сўзшакл бўлиб, унда морф шакли тасодифан ўхшаш бўлган айнан бир морфологик хусусиятга эга бўлган минимал ҳодиса; алломорф бир морфеманинг турли фонетик таркибга эга бўлган морфлари йиғиндиси; сўзшакл эса таркибий қисмлари орасида морфотактика орқали аниқланадиган алоқа мавжуд бўлган алломорфлар кетма-кетлиги ҳисобланади. Морфотактик қоидалар эса асос ва аффиксал гуруҳлар орасидаги ўзаро алоқани аниқлашга хизмат қилади.

6. Морфологик таҳлил стемматизация, лемматизация, граммема-лаш каби асосий блоклардан ташкил топади. Бунда стемматизация қидириладиган сўзнинг асосини топиш; лемматизация сўзшаклни леммага келтириш жараёни бўлиб, у флексив тиллар учун аҳамиятли, чунки агглютинатив тилларда сўзнинг “стем”и лексемага тенг бўлади. Фақат флексияга учраган сўзшаклда лемма ва стемнинг мос келмаслик ҳолати кузатилади. Таҳлил сифатини оширишнинг яна бир омили позицион ва нопозицион морфемаларни аниқлашдан иборат. Позицион морфема гапдаги позициясидан қатъи назар бир семантик гуруҳга мансуб бўлади. Нопозицион морфема гапдаги позициясига қараб турли семантик гуруҳга мансуб бўлади. Маълумотлар базасида позицион морфемалар рўйхатини шакллантириш, уларнинг локализациясини моделлаштириш морфологик таҳлил сифатини оширади.

7. Грамматик категорияга муносабати жиҳатидан сўз шакллари *категориал ва нокатегориал шаклларга* бўлинади. Категориал шакл маълум бир сўз туркумига хос грамматик шакллар системасига бирлашувчи, грамматик категория ҳосил қилувчи шакл, улар граммема сифатида белгиланади. Нокатегориал шакл ҳам маълум грамматик маънони англатади, аммо ўзаро бир системага бирлашмайди, маълум даражада мустақилликка эга бўлади, улар квазиграммема саналади. МАда сўз ясовчи шакл дериватема саналади. Ўзбек тили морфоанализи (1) эҳтимолий ўзакни қидириш (морфологик анализатор сўров ойнасига тушган сўзшакл асослар луғатида эҳтимолий бўлинадиган икки қисмга ажратилади); (2) асослар луғатидан эҳтимолий мос келувчи асосни қидириш; (3) маълумотлар омборидаги ахборот асосида топилган ўзакнинг морфонологик типини аниқлаш; (4) ўнг томондаги қисмларни таҳлил қилиш (аффикслар гуруҳи аниқланади) каби босқичлардан иборат бўлади.

8. Грамматик воситанинг сўздаги тартиби маълум қонуниятга эга, сўздаги қуршови эса унинг вазифасини белгилайди. Шакли бир хил, сўздаги қуршови ҳар хил бўлган грамматик шакллар морфотаҳлил учун алоҳида

моделлаштирилиши лозим. Шакли, сўздаги қуршови бир хил бўлган грамматик шаклни аниқлашга морфоанализ билан ҳам эришиб бўлмайди, улар семантик/синтактик автоматик ишловдан ўтиши лозим. Сўзда луғавий шакл ясовчи морфема қуршовини моделлаштириш морфотаҳлилда унинг сўз ясовчи, синтактик шакл ясовчидан фарқланишини таъминлайди. Морфоанализатор ясама сўз, ясовчи қисм ва ясалиш асосини аниқлаши учун лингвистик таъминотда ясама сўз структураси ҳақида маълумот бўлиши лозим. Сўз ясалиш моделлари асосида анализатор ўнгдан чапга қараб уч нарсани: сўзнинг асоси, ясалиш асоси ва сўз ясовчи аффиксни аниқлайди. Модел морфоанализаторнинг шаклни таниши, асос туркумини аниқлаши, асосга қўшилган морфемалар тартиби ҳамда бу тартибнинг маънога таъсирини таҳлил қилиши нуктаи назаридан ишлаб чиқилади.

9. Адабий тил нормасидан четлашиш ҳолати қаҳрамоннинг индивидуал лексикаси, характерини яратишга хизмат қилади. Адабий тил учун типик бўлмаган, аммо нутқда кўп учрайдиган ҳодисалар ҳам морфоанализатор базасида эквивалент сифатида қайд этилиши лозим, чунки фойдаланувчи мақсади билан бадиий асар парчаси, оғзаки нутқ (баъзан шева) материали ҳам морфологик таҳлилга тортилиши мумкин. Қўшимча морфемалардаги истисно “истиснолар луғати”дан жой олиши талаб қилинади. Бундай шакллар рўйхатини тузишда адабий меъёр билан бирга номеъёрий, аммо кенг ишлатиладиган шакллар инобатга олиниши зарур.

10. Морфоанализаторнинг ўзбек тили луғавий шакл ясовчи морфемалар қуршовини аниқлаши учун 9 та инвариант модел асосида 39 та субмодел; сўз ясовчи қўшимчалар қуршовининг автоматик таҳлили учун 1 та инвариант модел асосида 7 та субмодел, 60 та субмодел кўриниши ишлаб чиқилди. Мураккаб сўз ясалиши автоматик таҳлили учун, асосан, инвариант ва субмоделлар ишлаб чиқилди. Жумладан, ишимизда ўзакка иккита сўз ясовчи морфема қўшилишини фарқлаш учун битта инвариант асосида 7 субмодел, унинг 5 та кўриниши; ўзакка учта, тўртта сўз ясовчи морфема қўшилишини фарқлаш учун битта инвариант модел аниқланди.

**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**

ХАМРАЕВА ШАХЛО МИРДЖОНОВНА

**ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ МОРФОЛОГИЧЕСКОГО
АНАЛИЗАТОРА УЗБЕКСКОГО ЯЗЫКА**

10.00.11 – Теория языка, прикладная и компьютерная лингвистика

АВТОРЕФЕРАТ

ДОКТОРСКОЙ ДИССЕРТАЦИИ (DSc) ПО ФИЛОЛОГИЧЕСКИМ НАУКАМ

Фергана - 2021

Тема диссертации на соискание ученой степени доктора наук (DSc) зарегистрирована за номером B2020.4.DSc/Fil227 в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан.

Диссертация выполнена Бухарском государственном университете.
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

Искандарова Шарифа Мадалиевна
доктор филологических наук, профессор

Шахабитдинова Шохиди Хашимовна
доктор филологических наук, профессор

Ведущая организация: Самаркандский государственный университет

Защита диссертации состоится 28 06 2021 года в 9.00 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@mail.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Ферганского государственного университета (зарегистрирована за номером 412). (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@mail.uz).

Автореферат диссертации разослан 14 06 2021 года

(Реестр протокола рассылки № 22 06 2021 года).



А.А.Касимов
Председатель Научного совета по
присуждению ученых степеней, доктор
филол. наук, профессор

М.Т.Закиров
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 (Автореферат докторской диссертации)

Актуальность и востребованность рассматриваемой проблемы в выбранной научной отрасли. В мировой компьютерной лингвистике автоматический морфологический анализ изучается с 1950-х годов, были предприняты попытки применить его в практике обработки естественного языка, математической лингвистики и машинного перевода. Разработка морфологических анализаторов тюркских языков началась в 60-х годах прошлого века. Искусственный интеллект расширил свои возможности в области автоматического перевода, компьютерного анализа, редактирования, тезауруса, электронного словаря. Эти обновления в науке проложили путь к появлению перспективных научных направлений, связанных с внедрением информационных технологий в лингвистику. Так как, дело дошло до необходимости изучения общих принципов морфологического анализа в информационно-поисковой системе, корпусной лингвистике и машинном переводе, разработки системы создания формального языка и математики для лингвистического обеспечения систем анализа.

В мировой практической лингвистике в конце XX века на повестке дня изучаются такие вопросы как, улучшение качества автоматического перевода, лингвистическое моделирование языка, теория лемматизации слов в пределах определённого языка. В итоге, в информационно-поисковой системе были разработаны следующие приложения как, активно используемый стеммер, орфографический корректор. Следовательно, создание языковых корпусов, их лингвистическая аннотация, программы автоматической обработки текста – морфоанализатор, лемматизатор, стеммер, парсер, разработка орфокорректоров превратилось в одну из важнейших задач в корпусной лингвистике.

В узбекской компьютерной лингвистике за годы независимости был проведен ряд исследований по корпусной лингвистике, синтезаторам речи, автоматическому переводу, искусственному интеллекту для понимания и обработки узбекского языка, но, к сожалению, инструменты автоматического морфологического анализа – создание морфоанализатора узбекского языка в монографическом плане в целом не изучались. Следовательно, «... сохранить безупречность государственного языка, обогатить его и повысить речевую культуру населения; обеспечение активной интеграции государственного языка в современные информационные технологии и коммуникации»¹ подразумевает практическую эффективность любого теоретического исследования. Соответственно, такие вопросы, как языковой корпус, электронный переводчик, тезаурус, орфокорректор, создание исходных средств автоматической обработки и их лингвистическое обеспечение, имеют большое значение для достижения обработки узбекского языка в современных информационных технологиях.

¹Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг 2020 йил 202 октябрдаги «Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида»ги ПФ-6084-сон фармони // <https://lex.uz/docs/5058351>

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.

Соответствие диссертации приоритетам развития науки и технологий Республики Узбекистан. Исследование проводилось в соответствии с приоритетным направлением развития науки и технологий республики Узбекистан. 1. «Духовно-нравственное и культурное развитие демократического и правового общества, формирование инновационной экономики».

Обзор зарубежных исследований по теме диссертации². Научные исследования, направленные на изучение лингвистического обеспечения 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

² Диссертация мавзуси бўйича хорижий илмий тадқиқотлар шарҳи google.scholar.com; www.bu.edu; americanenglish.state.gov; micros.uz; www.uv.es; www.zora.uzh.ch; cleee@snu.ac.kr; www.u-tokyo.ac.jp/en; www.ras.ru; adwww.ecu.edu.au; unice.fr/en; www.sophia.ac.jp/eng; www.msu.ru; www.vsu.bu; www.education.ua/universities; www.kaznu.kz; www.navoiy-uni.uz ва бошқа манбалар асосида амалга оширилди.

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-Хоразми и его филиалы, Бухарский государственный университет, Термезский государственный университет (Узбекистан).

В мировой лингвистической информационно-поисковой системе (ИПС) были получены следующие научные результаты в результате исследования лингвистического обеспечения, технологий и инструментов автоматической обработки текста, программ автоматического анализа/синтеза текста, морфологических анализаторов: описывается лингвистическое обеспечение автоматизированных информационно-поисковых систем, определяются основные и вспомогательные средства поиска информации лингвистической ИПС, мультибазовые языковые средства (Cambridge University, Великобритания; University of Zurich, Швейцария); изучены особенности автоматической обработки естественного языка (анализ/синтез языковых единиц) (University of New Jersey, University of California, США; Boston University, New York University, США; Universitat de Valencia, Испания); выявлены способы использования трандюссеров при компьютерной обработке естественного языка (Courant Institute of Mathematical Sciences, New York, США; Pereira – San Francisco, USA, США); изучены технологии автоматической обработки текста, разработаны инструментарии и средства (University of Tokyo, Япония, Australian National University, Австралия, Oxford University Language Centre, Великобритания); разработаны теоретические основы морфологического анализа, его общее описание, принципы работы, грамматическая категория языка, описание лингвистических модулей для программы 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); модели и программы для автоматической морфологической и синтаксической обработки текста Харьковский государственный педагогический университет (Украина), Bakı Dövlət universiteti (Азербайджан); в компьютерной лингвистике тюркских языков разработаны задачи компьютерной обработки естественного языка, морфологического анализатора, языкового корпуса, морфоанализатора, средств синтеза голосовой речи, электронных переводчиков Институт прикладной семиотики Академии наук Республики Татарстан (Республика

Татарстан, Россия), Казанский федеральный университет (Республика Татарстан, Россия), 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 Национальный университет Узбекистана имени Мирзо Улугбека, Ташкентский государственный университет узбекского языка и литературы имени Алишера Навои, Ташкентский университет информационных технологий имени Мухаммада аль-Хоразми и его филиалы, Бухарский государственный университет, Термезский государственный университет (Узбекистан).

В мировой лингвистике исследования ведутся по приоритетным направлениям таким как, программы анализа информационно-поисковых систем, языковые корпуса для облегчения обработки в современных инструментах информационных технологий, электронный переводчик, тезаурус, орфографический корректор, первые средства автоматической обработки (лингвистические анализаторы: лемматизатор, морфоанализатор, стеммер, парсер, орфокоорректор), также создание их лингвистического обеспечения.

Степень изученности проблемы. Изучение теоретико-методологических проблем лингвистического обеспечения в мировой лингвистике началось в 60-х годах прошлого века: было признано, что компьютер без информационно-поискового языка остаётся машиной, перешедшей из «механического» этапа в «логико-лингвистический» этап, то есть, переходит в этап информационно-поисковых систем³. В трудах А.Кейлера, К.В.Линдена, Н.Варда⁴, М.А.Мохри⁴ освещены проблемы обработки естественного языка в мировой компьютерной лингвистике. Описание технологий автоматической обработки текста, инструментариев были приведены в трудах В.Ахо Альфреда, С.Лам Моники, С.Рави, Д.Ульмана Джеффри⁶, теоретические основы морфологического анализа описываются в трудах С.Ю.Толдовой, А.А.Бонч-Осмоловской⁷, Т.Садикова⁸. В

³ Соколов А.В. Автоматизация библиографического поиска / А.В.Соколов. – М.: Книга, 1981. – 167 с. – С.91.

⁴ Speech and Language Processing. Daniel S. Jurafsky and James H. Martin. Contributing writers: Andrew Kehler, Keith Vander Linden, Nigel Ward 2000y. Prentice Hall, Englewood Cliffs, New Jersey 07632. pages: 950.

⁵ MohriM.A. Finite-state transducers in language and speech processing. Computational Linguistics. – С. 269-312.

⁶ Ахо Альфред В., Лам Моника С, Сети Рави, Ульман Джеффри Д. Компиляторы: принципы, технологии и инструментарий, 2-е изд.: Пер. с англ. – М.: ООО “И.Д. Вильямс”, 2008. – 1184 с.

⁷ Толдова С.Ю., Бонч-Осмоловская А.А. Автоматический морфологический анализ // Фонд знаний «Ломоносов». М., 2011. www.lomonosov-fund.ru/enc/ru/encyclopedia:0127430

⁸ Садыков Т., Кочконбаева Б. Об оптимизации алгоритма морфологического анализа // Шестая Международная конференция по компьютерной обработке тюркских языков «Turklang-2018». (Труды конференции) – Ташкент, 2018. – 320 с.

компьютерной лингвистике русского языка стоит отметить работы О.В.Дерезы, Д.А.Каютенко, А.С.Феногеновой, М.Хакимова по морфологическому анализатору, его общему описанию, принципам функционирования, грамматической категории языка, описанию лингвистических модулей для программы⁹. Подробное описание мы можем увидеть в исследованиях В.Г.Митрофанова в работе «Описание лингвистического обеспечения автоматизированных информационно-поисковых систем» и в трудах В.В.Пономарёва, Л.Н.Пирумов «Вопросы разработки основных и вспомогательных средств ИПС, средств мультибазового языка». О важности лингвистического обеспечения мы можем узнать в трудах А.Б.Антопольского, Ю.М.Арского, Г.Г.Белоногова, Б.А.Кузнецова, В.Н.Белоозерова, Р.Р.Мдивани, Б.А.Глинского, А.И.Михайлова, В.А.Мишина, В.В.Морозова¹⁰ Особенности морфоанализа изучал Д.Рудольф¹¹, описание практического значения морфоанализа провёл А.Ермаков, модели и программы автоматической морфологической и синтаксической обработки текста исследовал И.М.Ножов¹². Проведен ряд работ по информационному поиску, машинному переводу, потребности обращения к грамматико-морфологическому, синтаксическому анализу слов при автоматическом редактировании, оптимизации морфологического анализа.

В тюркской компьютерной лингвистике детально изучена компьютерная обработка естественного языка, в том числе построение морфоанализаторов. Среди таких работ мы можем отметить труды посвящённые вопросам создания морфологического анализатора А.В.Дыбо,

⁹ Дережа О.В., Каютенко Д.А., Феногенова А.С. Автоматический морфологический анализ для русского языка сравнительный анализ систем // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент, 2018.; Хакимов М. Логико-лингвистические модели русского языка // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент, 2018. – 320 с.

¹⁰ Антопольский А.Б. Лингвистическое обеспечение АСНТИ: структура и проблема совместимости // НТИ. Сер.2, 1983. – №7. – С.17-25.; Арский Ю.М. Информационные ресурсы для устойчивого развития общества / Ю.М.Арский, А.И.Черный // Междунар. форум по информ. и док. – 2000. – Т. 25, № 1. –С. 3-9.; Белоногов Г.Г. Автоматизированная обработка научно-технической информации. Лингвистические аспекты / Г.Г. Белоногов, Б.А. Кузнецов, А.П. Новоселов // Итоги науки и техники. Сер. Информатика / ВИНТИ. –1984. –Т.8. – 316.; Белоозеров В.Н. Ведение и совершенствование Государственного рубрикатора НТИ (ГРНТИ) / В.Н. Белоозеров, В.И. Федосимов, Г.А. Кривенцова // Информ. продукты, процессы и технологии: НТИ-96: Материалы конф. – М., 1996. – С. 60-69.; Мдивани Р.Р. Лингвистическое обеспечение автоматизированной информационной системы по общественным наукам: состав и структура / Р.Р.Мдивани, В.А.Глинский // Вопр. информ. теории и практики. 1985. –№52. –С. 37-55.; Михайлов А.И. Проблемы лингвистического обеспечения АСНТИ / А.И. Михайлов // Вопр. информ. теории и практики. 1984.–№ 51. – С. 5-11.; Мишин В.А. Лингвистическое обеспечение библиотечных автоматизированных систем / В.А. Мишин, Л. Санчес-Мехидо // Сов. библиотековедение. 1981. – № 2. – С. 84-93.; Морозов В.В. Лингвистическое обеспечение автоматизированных систем научно-технической информации / В.В. Морозов, Б.Р. Певзнер. –М.: ИПКИР, 1987. – 76 с.

¹¹ Rodolfe D. Computational Linguistic Text Processing: Lexicon, Grammar, Parsing and Anaphora Resolution. NovaSciencePublishers, Inc. – NewYork, 2008. – Р. 4-5.; Ермаков А. Морфологический анализатор - основа поисковых систем // <https://www.kv.by/archive/index2004154301.htm>

¹² Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы): диссертация канд. наук. – Москва, 2003. – 190 с.

А.В.Шеймовича¹³, Ж.Сулейманова¹⁴, Г.Эрйигита, Е.Адали,
А.Р.Гатиатуллина, А.М.Баширова¹⁵, П.В.Желтова¹⁶, Н.А.Исраиловой,
П.С.Бакасовой¹⁷, Н.А.Леонтьева¹⁸, В.В.Кукановой, А.Ю.Каджиева¹⁹.

Система морфологического анализа тюркских языков и описание их моделей, широта исследований, посвященных анализу, укрепили теоретические основы диссертации. Также были реализованы теоретические идеи разработки морфоанализаторов, разработаны морфоанализаторы европейского, русского и тюркских языков. В узбекском языкознании был проведен ряд исследований по компьютерной лингвистике, лексикографической обработке текста и лингвостатистическому анализу. В качестве выше указанных трудов, мы можем привести пример труды следующих исследователей, таких как М.А.Махмудов, М.Айимбетов, С.Каримов, Г.Жуманазарова, А.Бабанаров, Д.Ўринбаева, А.Норов, С.Мухаммедов, Н.Абдурахманова и другие²⁰. Разработка морфологического

¹³ Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов хакасского и древнетюркского языков / Научное обозрение саяно-алтая рецензируемый научный журнал Номер 2(08), 2014 (Периодичность – 2 раза в год. серия: Филология.Выпуск 2). С. 9-31.

¹⁴ Сулейманов Д.Ш., Гильмуллин Р.А., Гатауллин Р.Р. Морфологический анализатор татарского языка на основе двухуровневой модели морфологии / 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.; Сулейманов Д.Ш., Гатиатуллин А.Р. Модель татарской аффиксальной морфемы и ее реализация // Серия: Интеллект. Язык. Компьютер. – Вып.4. – Казань: Фэн. – 1996. – 113-127.

¹⁵ Eryigit G., Adali E. An affix stripping morphological analyzer for Turkish \ Proceedings of IASTED International Conference ARTIFICIAL INTELLIGENCE AND APPLICATIONS. February 16-18, 2004, Innsbruck, Austria. – P. 299-304.; Гатиатуллин А.Р., Баширов А.М. Морфологический анализатор тюркских словоформ на базе структурно-функциональной модели тюркской морфемы // 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.

¹⁶ Желтов П.В. Разработка морфологического анализатора чувашского 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.

¹⁷ 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.

¹⁸ Леонтьев Н.А. Морфологический анализатор якутского языка // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент, 2018. – 320 с. – С. 276-279.

¹⁹ Куканова В.В., Каджиев А.Ю. Алгоритм работы морфологического парсера калмыцкого языка // В сборнике: Писменото наследство и информационните технологии. E!Manuscript–2014 Материали от V международна научна конференция. Отговорни редактори В.А.Баранов, В.Желязкова, А.М.Лаврентьев. 2014. – С. 116-119.

²⁰ 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., Каршиев А., Исраилова Г. Абдулла Қаҳхор асарлари тилининг лугати. Алфавитли лугат. Частотали лугат. Терс лугат. – 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 лингвопoэтикаси: Фил. фан. док. дисс...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. – Самарқанд, 2019. – 74 б.

анализатора узбекского языка не была предметом специальных исследований, но в некоторых исследованиях были даны комментарии по определенным аспектам проблемы. В частности, в монографическом плане изучалось создание лингвистического программного обеспечения для компьютерных программ на основе глаголов действия²¹, лингвистическое обеспечение узбекско-английского машинного перевода²² принципы создания узбекского авторского корпуса²³, проблемы графематического анализа узбекских языковых единиц²⁴, принципы создания лингвистической базы языкового корпуса. Также опубликовано ряд статей о лингвистических основах семантического тегирования узбекских назывных единиц²⁵ проблемах создания узбекско-английского параллельного корпуса²⁶, автоматическом анализе в узбекском языке имён существительных, формальных моделях словообразования²⁷. Изучены программные аспекты автоматизации морфологического анализа имён существительных и прилагательных в узбекском языке²⁸. Данные исследования актуальны, так как изучают проблемы компьютерной лингвистики, но вопрос автоматического морфологического анализа узбекского языка – морфоанализатора и принципов создания его лингвистического обеспечения не стоит на повестке дня. В процессе подготовки диссертации указанные научные исследования были подробно изучены, при необходимости обработаны и использованы в исследованиях.

Соответствие исследования исследовательским планам высшего учебного заведения или исследовательского учреждения, в котором была выполнена диссертация. Диссертация выполнена в рамках проекта

²¹ Мухаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. – Тошкент, 2006.;

²² 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.

²³ Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2018. – 250 б.

²⁴ 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.

²⁵ Akhmedova D.B., Mengliev B.R. Semantic Tag Categories in Corpus Linguistics: Experience and Examination International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-3S, October 2019. – P. 208-212.

²⁶ 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 ngineering and Advanced Technology (IJEAT). ISSN: 2249 – 8958, Volume-8, Issue-5S3 July 2019. – P. 388-391.

²⁷ 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.

²⁸ Болтаев Т.Б., Ибрагимов С.И. О проекте программной системы морфологического анализа узбекского языка // <http://buxdu.uz/index.php/uz/>; Болтаев Т.Б., Кузьминов Т.В., Поттосин И.В. О структурном конструировании программ и инструментах его поддержки // Среда программирования: методы и инструменты. Новосибирск, 1992, C22-37.; The Structured Constructing as a Discipline of Safe Programming and Instruments Supporting It / Aniskov M.I., Boltaev T.B., Kochetov D.V. at al // Instrumental Congress on Computer Systems and Applied Mathematics CSAM'93. St-Petersburg. July 19-23.

«Создание узбекского образовательного корпуса» под номером АМ-ФЗ-201908172 предназначенному выполнению в 2020-2021 годы в Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои.

Целью диссертации раскрыть основы лингвистического обеспечения морфологического анализатора узбекского языка.

Задачи диссертации: общее описание языковых средств, их значение в создании морфологического анализатора и языкового корпуса;

автоматический морфологический анализ языковых единиц/изучение средств процесса синтеза;

исследование структуры, состава и функций лингвистического обеспечения;

общие принципы создания морфологического анализатора и его значение в информационно-поисковой системе;

исследование общих и различных аспектов морфологических анализаторов тюркских языков;

определение единиц, структуры, состава и анализ морфологического анализатора узбекского языка;

лингвистическое моделирование словообразовательных морфем, синтаксического формообразования, словообразовательных морфем, а также дополнительных формообразующих частиц.

В качестве **объекта исследования** было выбрано лингвистическое обеспечение морфологического анализатора узбекского языка.

Предмет исследования являются основы моделирования и систематизации единиц лингвистического обеспечения.

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

Научная новизна исследования заключается в следующем:

лингвистическое обеспечение информационно-поисковой системы на узбекском языке основано на функциональном аспекте таких анализаторов, как орфо корректор, лемматизатор, морфоанализатор, парсер путем определения состава и функций;

доказано, что морфологический анализатор в информационно-поисковой системе является основным инструментом лингвистической аннотации текста и автоматической обработки естественного языка при машинном переводе;

установлено, что функциональные возможности морфологических анализаторов, разработанных в тюркских языках, различны с точки зрения лингвистического обеспечения, автоматическая обработка языка является общей с точки зрения методов морфологического анализа;

разработаны комбинаторные и позиционные правила морфотактики для слов, отличающихся от общего морфолексикона за счёт орфографических изменений в узбекском языке;

лингвистическое моделирование программного алгоритма словообразования, словаря-формообразования, синтаксического формообразования морфем для автоматического морфологического анализа единиц узбекского языка, а также словосочетательных позиций дополнительных формообразующих нагрузок.

Практические результаты исследования следующие:

проанализированы средства автоматического процесса морфологического анализа/синтеза языковых единиц, определены общие принципы создания морфологического анализатора, обоснована его значимость в информационно-поисковой системе, языковом корпусе, машинном переводе и компьютерной лексикографии;

показаны общие и различные аспекты морфологических анализаторов тюркских языков, описаны единицы морфоанализатора узбекского языка, структура, состав, сформирован список единиц анализа в исследовательском приложении;

было внесено предложение о разработке морфологического анализатора узбекского языка и его лингвистического обеспечения;

лингвистически смоделированы словообразовательные, формаобразовательные, словоизменительные морфемы для морфологического анализатора и окружение частиц в форме аффиксов представлений 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 принципов морфологического анализатора узбекского языка.

Научная и практическая значимость результатов исследования. Исследование играет теоретическую роль в разработке теоретических основ для создания морфологического анализатора узбекского языка и его лингвистического обеспечения, в создании исследований в области компьютерной лингвистики. Практическая значимость исследования объясняется тем, что в преподавании прикладной лингвистики, компьютерной лингвистики корпусная лингвистика может использоваться как источник при разработке программ, планов и описаний тем, а также при разработке морфологического анализатора узбекского языка, орфо корректора, морфологической разметки различных типов корпуса на узбекском языке.

Внедрение результатов исследований.

По результатам лингвистического моделирования лингвистического обеспечения узбекского морфологического анализатора, общих принципов работы морфологического анализатора, значений морфологического анализатора в информационно-поисковой системе, единиц, структур и состава морфологического анализатора узбекского языка, единиц анализа, правил морфотактики морфологического анализатора, синтаксического

словообразователя (словоизменительного аффикса), формаобразователя и среды частиц в форме аффиксов:

по результатам лингвистического моделирования структуры, состава и единиц анализа, словообразования, лексического формообразования, синтаксического формообразования и моделирования среды частиц в форме аффиксов разработаны единицы морфоанализатора узбекского языка и использованы в фундаментально научном проекте по теме «Функциональное словообразование в современном каракалпакском языке» ФА-Ф1-Г003 в Каракалпакском филиале Академии наук Республики Узбекистан Каракалпакского научно-исследовательского гуманитарного института (справка № 89-03-4852 Министерства высшего и среднего специального образования от 20 ноября 2020 года). В результате были усовершенствованы функциональные модели словообразования в современном каракалпакском языке; каракалпакские формообразующие морфемы, принципы автоматической среды омонимии частиц в форме аффиксов;

выводы о важности использования единиц узбекского языка в информационно-поисковой системе, роли информационно-поисковой системы в формировании языковой культуры использованы в проекте в рамках государственных программ фундаментальных исследований «ОТФ-1-18/разработка методов и методологии формирования массовой языковой культуры» при Андижанском государственном университете намеченным на 2017-2020 гг. (справка № 89-03-4852 Министерства высшего и среднего специального образования от 20 ноября 2020 года). В результате разработаны принципы широкого использования единиц узбекского языка в информационно-поисковой системе при разработке методов и методологии формирования массовой языковой культуры, обоснована роль информационно-поисковой системы в формировании языковой культуры;

результаты лексических формообразователей и синтаксические морфемы, образующие омонимы с другими грамматическими формами также моделирования словесной среды с целью определения омонимических и полифункциональных свойств словообразовательной морфемы в контексте выполнены в Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои под номером I-ОТ-2019-42 в фундаментально-научном проекте «Создание электронного поэтического словаря узбекского и английского языков (изображение человека, персонажа, природы и национальных символов)» (справка № 89-03-4852 Министерства высшего и среднего специального образования от 20 ноября 2020 года). В результате в узбекском и английском языках появилась возможность автоматически определять контекстную омонимию слов, выражающих образ человека, персонажа, природы и национальных символов;

по результатам лингвистического обеспечения морфологического анализатора узбекского языка, общих принципов работы морфологического анализатора, единиц, структур и состава морфологического анализатора узбекского языка, единиц анализа, правил морфотактики была разработана ЭВМ программа для «Национального и авторского корпуса узбекского

языка» (Свидетельство об официальной регистрации программы для электронных-вычислительных машин. Агенство по интеллектуальной собственности республики Узбекистан № DGU 05735). В результате разработано лингвистическое обеспечение национального и авторского корпуса узбекского языка, доработаны единицы языкового корпуса, аналитические единицы, система тегов;

сделаны выводы о среде суффиксов для определения значения словообразовательных суффиксов, об изменении значения словообразовательных суффиксов в зависимости от положения в слове и по данным выводам были использованы в учебнике «Современный узбекский язык» 13-параграфа «Лексема производных слов», где описываются структуры образования лингвистических производных слов и чисто произнесенных производных слов в лексике узбекского языка. Данный учебник предназначен для студентов бакалавриат по направлению 5220100 – Филология и преподавание языков (справка № 89-03-4852 Министерства высшего и среднего специального образования от 20 ноября 2020 года). В результате улучшено описание структуры образования лингвистических производных слов и чисто разговорных производных слов в лексике узбекского языка. При преподавании дисциплины «Корпусная лингвистика» магистрам по направлению 5A120106 – Компьютерная лингвистика, на лекциях также при выполнении практических заданий магистранты пользуются руководством для практических работ «Корпусная лингвистика: создание корпуса и пользование им» также монографией «Лингвистические основы создания авторского корпуса узбекского языка» опубликованные автором (справка № 89-03-4852 Министерства высшего и среднего специального образования от 20 ноября 2020 года). В результате был расширен круг литературы и источников информации при преподавании предмета «Корпусная лингвистика: создание корпуса и пользование им»;

выводы о важности автоматических переводчиков, конвертере кириллица-латиница, компьютерного анализа, редактирования, языковых корпусов, электронных словарей и тезаурусов, доступных в Интернете, в повышении общественной грамотности, формировании языковой культуры среди молодежи и использовании узбекского языка в современных информационных системах были использованы при трансляции передачи «Ассалом, Бухоро» телеканала «Бухоро» Бухарской областной телерадиокомпании (справка № 1/296 от 12 ноября 2020 года Бухарской областной телерадиокомпании). В результате было улучшено содержание пропагандистской работы по воспитанию высокого гармонично развитого поколения, формированию языковой культуры у молодежи и использованию узбекского языка в современной информационной системе.

Апробация результатов исследования. Результаты данного исследования обсуждались на 12 международных и 18 национальных конференциях и доступны в следующих профилях: orcid.org/0000-0002-5429-4708, www.researchgate.net/profile/Shahlo_Khamroeva, [https:// scholar.google.com/ citations](https://scholar.google.com/citations)

Публикация результатов исследования. По теме диссертации опубликовано 54 научных работы, в том числе 1 монографии, 11 научных статей в 7 научных изданиях, рекомендованных к публикации основных научных результатов докторских диссертаций ВАК при Кабинете Министров Республики Узбекистан, из них 4 в зарубежных журналах.

Объем и структура диссертации. Диссертация состоит из введения, четырех глав, заключения, списка использованной литературы и приложения. Размер 264 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

В **вводной части** диссертации обоснована актуальность и востребованность темы диссертации, охарактеризованы цели и задачи, объект и предмет исследования, соответствие работы приоритетным направлениям развития науки и технологий республики Узбекистан, научная новизна и практические результаты. Освещены практическая и научная значимость полученных результатов, степень изученности темы и методы исследования. Предоставлена информация о внедренности в практику результатов исследования, опубликованных работах и содержании диссертации.

Первая глава диссертации озаглавлена **«Роль лингвистического обеспечения в информационно-поисковых системах»**. Первый раздел главы *«Общее описание средств лингвистического обеспечения, морфологического анализатора, его роль в формировании языковых корпусов»* описывает лингвистическое обеспечение, роль лингвистического обеспечения в информационно-поисковых системах. Изучение теоретико-методологических проблем лингвистического обеспечения в мировой лингвистике началось в 60-х годах прошлого века. О важности лингвистического обеспечения мы можем прочитать в трудах²⁹ А.Б.Антопольского, Ю.М.Арского, Г.Г.Белоногова, Б.А.Кузнецова, В.Н.Белоозерова, Р.Р.Мдивани, Б.А.Глинского, А.И.Михайлова, В.А.Мишина, В.В.Морозова.

Лингвистическое обеспечение – это набор инструментов, используемых для автоматической обработки текстов на естественном языке

²⁹ Антопольский А.Б. Лингвистическое обеспечение АСНТИ: структура и проблема совместимости // НТИ. Сер.2, 1983. – №7. – С.17-25.; Арский Ю.М. Информационные ресурсы для устойчивого развития общества / Ю.М.Арский, А.И.Черный // Междунар. форум по информ. и док. – 2000. – Т. 25, № 1. – С. 3-9.; Белоногов Г.Г. Автоматизированная обработка научно-технической информации. Лингвистические аспекты / Г.Г.Белоногов, Б.А.Кузнецов, А.П.Новоселов // Итоги науки и техники. Сер. Информатика / ВИНТИ. – 1984. – Т.8. – С.316.; Белоозеров В.Н. Ведение и совершенствование Государственного рубрикатора НТИ (ГРНТИ) / В.Н.Белоозеров, В.И.Федосимов, Г.А.Кривенцова // Информ. продукты, процессы и технологии: НТИ-96: Материалы конф. – М., 1996. – С.60-69.; Мдивани Р.Р. Лингвистическое обеспечение автоматизированной информационной системы по общественным наукам: состав и структура / Р.Р.Мдивани, В.А.Глинский // Вопр. информ. теории и практики. 1985. – №52. – С. 37-55.; Михайлов А.И. Проблемы лингвистического обеспечения АСНТИ / А.И.Михайлов // Вопр. информ. теории и практики. 1984. – № 51. – С. 5-11.; Мишин В.А. Лингвистическое обеспечение библиотечных автоматизированных систем / В.А. Мишин, Л.Санчес-Мехидо // Сов. библиотековедение. 1981. – № 2. – С. 84-93.; Морозов В.В. Лингвистическое обеспечение автоматизированных систем научно-технической информации / В.В.Морозов, Б.Р. Певзнер. – М.: ИПКИР, 1987. – 76 с.

(язык запросов и поиск), языковой процессор. Лингвистическое обеспечение – это набор символов³⁰, обеспечивающих адекватную работу языковых инструментов в конкретном социальном общении, языки, используемые в программном обеспечении информационных технологий, способ/средства их создания, комплекс применения и использования³¹. Таким образом, можно сделать вывод, что лингвистическое программное обеспечение – это набор программного обеспечения и данных, связанных с анализом, обработкой, хранением аудиоинформации, изображения и текста на естественном языке. Лингвистическое обеспечение информационно-поисковой системы включает в себя³²:

1. Информационно-поисковый язык (содержание документа и поисковый язык): алфавит и микросинтаксис ИПС; средства выражения контент-плана: словарь, тезаурус, вспомогательные словари; классификатор; грамматические средства ИПС; средства выражения информации на естественном языке.
2. Средства адаптации разных языков: таблица совместимости; график конверсии; протоколы.
3. Практические инструменты: методы индексации; алгоритм автоматической индексации; поисковый инструмент.
4. Критерии эквивалентности контента.
5. Структурированные языки и метаописания данных.
6. Вспомогательные инструменты для обработки текста на естественном языке.
7. Основные инструменты лингвистического обеспечения: создание электронного словаря; описание процесса работы с электронным словарем; документация лингвистического обеспечения³³.

На наш взгляд, лингвистическое обеспечение – это набор языковых инструментов, которые служат для создания формальных правил естественного языка с целью повышения эффективности автоматического анализа информации, используемой в автоматизированной поисковой системе. Оно состоит из структуры объекта информации, действий, выполняемых над понятием, используемого в его работе, формализованного представления на естественном языке, классификации и набора ее средств.

Во втором разделе главы, озаглавленном «*Морфологический анализ и синтез словоформ*», обсуждается идеальный лингвистический (лексический, морфологический, синтаксический, семантический) анализ текста или/и информации. Лучшие доступные информационно-поисковые системы выполняют морфологический анализ текстовой информации, а

³⁰ https://sociolinguistics.academic.ru/915/%D0%9B%D0%B8%D0%BD%D0%B3%D0%B2%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B5_%D0%BE%D0%B1%D0%B5%D1%81%D0%BF%D0%B5%D1%87%D0%B5%D0%BD%D0%B8%D0%B5.

³¹ <https://works.doklad.ru/view/IP7CrzcUzY.html>

³² https://studopedia.ru/6_11435_ponyatie-i-sostav-lingvisticheskogo-obespecheniya.html

³³ Пономарев В.В. Лингвистическое обеспечение и социолингвистическая специфика проблемы автоиндексационной актуализации информационных систем: автореф. диссер. канд. филол. наук. – Москва, 2005.

также синтаксический анализ предложения³⁴. Задача морфологического анализа заключается в том, чтобы автоматически определить, принадлежит ли каждое слово в тексте к части речи; чтобы различать, к какому лексико-грамматическому классу принадлежит слово. Синтаксический анализ текста позволяет выделить смысловые элементы предложения – принадлежность к группе, предикативной основе. Лексический анализ – это анализ текстовой информации в виде отдельных заголовков, предложений, слов, отражающих выражения национального языка. Форма речи, указывающая на специфику языка, определяет лексическое выражение. Языки программирования требуют дополнений к формальным языковым правилам. Синтаксис естественного языка работает по принципу «исключения подтверждают правило»³⁵. Грамматика – это описание методов построения предложений определенного языка; математическая система, определяющая язык. Она создана на основе формального описания грамматики, основанного на системе правил для языков программирования. В этом разделе также подчеркивается важность использования международных стандартов в лингвистике и программном обеспечении.

Третий раздел главы, озаглавленный «*Структура лингвистического обеспечения, состав и задачи*», описывает базу данных, которая является частью лингвистического обеспечения. Среди инструментов, основанных для разработки лингвистического обеспечения, мы можем добавить следующие³⁶:

1. Лингвистические процессоры (linguistic processors) – программа, автоматически обрабатывающая текстовую информацию и её алгоритм; процесс создания и применения языковых инструментов.

2. База лингвистических данных (linguistic data bank) хранит базу данных электронного словаря, важные записи и инструменты управления программным обеспечением, которые ими управляют.

На основании вышеизложенного можно сделать вывод, что лингвистическое программное обеспечение является необходимым компонентом программного обеспечения для машинного перевода, компьютерной лексикографии, корпусной лингвистики, процесса морфологического анализа.

II глава диссертации озаглавлена «Теоретические основы создания морфологического анализатора». Первый раздел главы «*Основные принципы создания морфоанализатора*» посвящен созданию морфологического анализатора (МА) и морфологической базы данных. Морфологический анализатор – это набор алгоритмов, предназначенных для сравнения словоформы и отдельных слов в словаре, чтобы предоставить грамматические особенности слова³⁷. Основная задача МА – определить

³⁴ Место и назначение лингвистического обеспечения в информационных системах. Понятие информационной системы // <https://gigabaza.ru/doc/116551-pall.html>

³⁵ Тот же источник.

³⁶ Тот же источник.

³⁷ http://www.solarix.ru/for_developers/docs/morphology_analyzer.shtml

вероятное грамматическое значение незнакомого слова. Если анализируемое слово не найдено в словаре, МА должен сделать возможные выводы о его грамматической роли. Для выполнения этой задачи требуется ряд информации, в частности набор правил анализа. Автоматический анализ и синтез текста – сложная операция: компьютер выполняет анализ текста на естественном языке на основе заданного алгоритма. В процессе автоматического анализа компьютер создает лексико-морфологическое, синтаксическое, смысловое представление текста на своем языке³⁸. Морфологический анализ текста позволяет выявить основную форму употребляемого в нем слова – лемму; определяются грамматические значения слова в контексте. Поскольку узбекский язык является агглютинативным, он часто эквивалентен основе леммы, как и во флективных языках, процесс морфоанализа не сложен. Следующим шагом будет определение морфологических признаков и принадлежности слова к категории. Суффиксы, которые выражают грамматическое значение словоформы, считаются образующими форму конкретной лексемы и ее формирователем как членом парадигмы. Парадигма всех суффиксов, которые могут быть присоединены к конкретной лексеме, создается в базе данных системы анализа. При этом также наблюдается грамматическая, лексико-морфологическая омонимия. В этом случае МА полагается на синтаксическую и семантическую особенность предложения при правильном присоединении морфологического знака к слову. Анализ в МА основан на двух принципах: «справа налево» и «слева направо»; значимая часть в конце слова – аффиксальная морфема; затем определяет наличие части в дополнение к суффиксу (справа) в словаре. Формирование морфологической базы данных – основной этап формирования МА. База данных – это широкий набор данных, который удовлетворяет потребности пользователя: в нем данные сортируются и хранятся в табличной форме. В настоящее время существуют такие системы, как SQL, MySQL, Oracle, Access, предназначенные для работы с базами данных.

Второй раздел главы, озаглавленный «*Значение морфоанализатора в информационно-поисковых системах*», представляет собой обзор программ обработки и анализа текста, их роль в информационно-поисковой системе. Следует отметить, что общей чертой изучаемых инструментов лингвистического анализа являются возможности графического, морфологического анализа. Уровень анализа этих программ варьируется в зависимости от признака языковой адаптации, некоторые из них имеют функцию семантического/синтаксического анализа, некоторые не имеют этой функции. На наш взгляд, включение нескольких функций в единую программу анализа увеличивает преимущество программы, поэтому при разработке МА узбекского языка необходимо уделять особое внимание одновременному охвату нескольких этапов анализа. Созданный

³⁸ Щипицина Л.Ю. Информационные технологии в лингвистике: учеб. пособие / Л.Ю.Щипицина. – М.: ФЛИНТА: Наука, 2013. – 128 с. – С. 43.

морфоанализатор узбекского языка превращается в многофункциональный анализатор, выполняющий различные лингвистические анализы.

Модуль графематического анализа текста для изучения особенностей программ морфологического анализа показал, что он состоит из различных тезаурусов и словарей; синтаксического словаря; модуля обнаружения и устранения омонимии; модуля начального семантического анализа; лингвистической поисковой системы (конкорданс); Например, RHRMorphy³⁹ предназначена для лемматизации; определение всех форм речи; объединение всей грамматической информации, относящейся к слову; также выполнения таких действий, как изменение словоформы в соответствии с заданной грамматической особенностью. Лемматизатор (Мультитран)⁴⁰ или русский морфологический анализатор MyStem⁴¹ также имеет свои преимущества. В качестве общей черты этих программ анализа важно подчеркнуть такие преимущества, как морфологический анализ, лемматизация, распознавание слов, грамматическая информация, относящаяся к слову, анализ нескольких языков. Отличие вышеуказанных МА состоит в том, что некоторые выполняют статистический, некоторые – семантический анализ. Наши наблюдения за МА славянского языка привели нас к выводу, что границы грамматической категории, грамматической формы, граммемы, квазиграммемы и дериватемы должны быть четко определены при разработке МА.

В третьем разделе главы, озаглавленном «*Общие и разные аспекты морфологических анализаторов тюркских языков*», обсуждаются описание морфоанализаторов, созданных в тюркских языках, исследования и эксперименты, комментарии к морфоанализаторам некоторых тюркских языков, исследования морфологического анализа узбекского языка.

Проблема автоматической обработки естественного языка широко изучалась в мировой лингвистике. В тюркской лингвистике также широко изучались вопросы компьютерной обработки естественного языка и морфологического анализатора. К таким работам относятся различные исследования А.В.Дыбо, А.В.Шеймовича⁴², А.Р.Гатиатуллина, А.М.Баширова⁴³, П.В.Желтова⁴⁴, Н.А.Исраиловой, П.С.Бакасовой⁴⁵.

³⁹ <https://novainfo.ru/article/3394>

⁴⁰ <https://www.multitran.com/c/m.exe?a=4&MessNum=8673&topic=2&l1=4&l2=2>

⁴¹ <https://yandex.ru/dev/mystem/>

⁴² Дыбо А.В., Шеймович А.В. Автоматический морфологический анализ для корпусов хакасского и древнетюркского языков / Научное обозрение саяно-алтая. Рецензируемый научный журнал Номер 2(08), 2014 (Периодичность – 2 раза в год, серия: Филология. Выпуск 2). С. 9-31.

⁴³ Гатиатуллин А.Р., Баширов А.М. Морфологический анализатор тюркских словоформ на базе структурно-функциональной модели тюркской морфемы // 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.

⁴⁴ Желтов П.В. Разработка морфологического анализатора чувашского 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.

⁴⁵ 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.

Хотя разработка морфологического анализатора узбекского языка не была предметом специальных исследований, в некоторых исследованиях были приведены комментарии по некоторым аспектам проблемы. Основываясь на наших наблюдениях над созданием морфоанализатора тюркских языков, мы пришли к выводу, что основателями МА являются следующие:

- 1) языковой словарь (словарь с описанием слова, относящегося к категории, также имеющий фонологические правила);
- 2) компьютерная модель, основанная на достаточном грамматическом описании словарного запаса (ориентированная на автоматический анализ языка);
- 3) свод правил выбора алломорфов, на которые распространяются фонетические правила добавления определенного суффикса к основе.

При создании морфоанализатора любого языка особое внимание уделяется особенностям языка. Специалисты описывают татарский язык как «семантически сложное, связанное, практически автоматическое языковое явление»⁴⁶. На основании изучения морфоанализатора татарского языка мы пришли к выводу, что язык имеет практическое значение в процессе поиска и обработки информации ИПС. Основная задача МА чувашского языка – определение морфемной структуры, морфологических особенностей чувашских слов для синтаксического, семантического анализа. База данных чувашского МА содержит список типов, описывающих морфологические признаки, словарь, справочную, базовую и суффиксную базу данных, хранение текста/слов, данные структуры рабочей модели, содержащие образец анализа слов, общие законы структуры слова, относящиеся к различным категориям слов⁴⁷. Аналогичный компонент имеет морфоанализатор кыргызского⁴⁸ и якутского⁴⁹ языков с некоторыми особенностями. Опыт при создании морфоанализаторов тюркских языков полезен при разработке морфоанализатора узбекского языка. Хотя морфоанализатор узбекского языка не исследовался в монографическом плане, есть труды⁵⁰ по некоторым вопросам морфологического автоматического анализа. Например, средства построения МА включают лексику, морфотактику, правила орфографии, единицы информационной структуры, которые формируют базовую семантику⁵¹. Изучение структуры, представленной в указанных выше работах, полезно при построении морфологического анализатора узбекского языка.

III Глава исследования озаглавлена «Лингвистические основы создания морфоанализатора узбекского языка». Первый раздел озаглавлен

⁴⁶ Сулейманов Д.Ш., Гильмуллин Р.А., Гатауллин Р.Р. Морфологический анализатор татарского языка на основе двухуровневой модели морфологии / 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.

⁴⁷ Желтов П.В. Кўрсатилган манба.

⁴⁸ Исраилова Н.А., Бакасова П.С. Кўрсатилган манба.

⁴⁹ Леонтьев Н.А. Кўрсатилган манба.

⁵⁰ Болтаев Т.Б., Ибрагимов С.И. О проекте программной системы морфологического анализа узбекского языка // <http://buxdu.uz/index.php/uz/>

⁵¹ Тот же источник.

«Единицы, структура и состав морфоанализатора узбекского языка». В источниках указываются стемминг автоматического анализа; анализ словоформ на основе словаря; анализ, основанный на логическом подходе; на основе таблицы показаны типы словарного анализа⁵². Автоматический морфологический анализ состоит из следующих блоков стемматизации, лемматизации, таггирования⁵³. Стемматизация – это процесс поиска основы искомого слова. Стем (основа словоформы) не должен быть равен грамматической основе словоформы, в котором выполняется поиск. Лемматизация – это процесс преобразования словоформы в лемму. Таггирования – это написание грамматической характеристики словоформы. Граммема (грамматическая характеристика) – простой морфологический показатель, указывающий на принадлежность словоформы к определенному морфологическому классу⁵⁴. В.А.Плунгян описывает граммему следующим образом⁵⁵: «Граммема – это грамматическое значение, которое является одним из элементов грамматической категории. Разные граммемы одной грамматической категории отличаются друг от друга и не могут быть выражены одновременно. Грамматический указатель – это тоже граммема. Граммема – это морфологическая форма, образующая морфологическую категорию. Следовательно, граммема – это грамматическое значение, это элемент грамматической категории, представленный грамматической формой.

Основными языковыми единицами, используемыми в МА, являются: морф – элементарный сегментный знак; алломорф – совокупность морф морфемы с одинаковым фонетическим содержанием⁵⁶; словоформа – последовательность алломорфов, в которой существует связь между компонентами, определяемая морфотактикой; морфема – наименьшая смысловая единица языка, составляющая слово, которая состоит из набора морфов с разными общими, схожими свойствами, выполняющих одну функцию; морфотактические правила – ключевой элемент, составляющий работу МА. Таким образом, для того, чтобы начать работу с морфоанализатором узбекского языка, необходимо разделить правила морфотактики, алломорфа, слова, морфемы.

Классификация основных морфем в лингвистическом обеспечении МА является важным процессом. В традиционной лингвистике известны принципы деления слов на части речи. Но в компьютерной лингвистике

⁵² Марчук Ю. Компьютерная лингвистика. – Москва, 2006. – С. 65.

⁵³ Формальные модели и программные инструменты компьютерной обработки татарского языка / Р.Р.Гатауллин, А.Р.Гатиатуллин, О.А.Невродова, Д.Р.Мухамедшин, Д.Ш.Сулейманов, Б.Э.Хакимов, А.Ф.Хусаинов. – Академия наук РТ, Институт прикладной семиотики АН РТ. – Казань: Изд-во Академии наук РТ, 2019. – 260 с. – С.39-40.

⁵⁴ Тот же источник.

⁵⁵ Плунгян В.А. Классификация морфологических значений // Общая морфология: Введение в проблематику: Учебное пособие. – Изд. 2-е, исправленное. – М.: Едиториал УРСС, 2003. – С. 107. – 384 с.

⁵⁶ Формальные модели и программные инструменты компьютерной обработки татарского языка / Р.Р.Гатауллин, А.Р.Гатиатуллин, О.А.Невродова, Д.Р.Мухамедшин, Д.Ш.Сулейманов, Б.Э.Хакимов, А.Ф.Хусаинов. – Академия наук РТ, Институт прикладной семиотики АН РТ. – Казань: Изд-во Академии наук РТ, 2019. – 260 с. – С.39-40.

мы наблюдаем другой подход к классификации слов. Слова классифицируются в базе данных «Словарь основ», поэтому в компьютерной лингвистике классификация слов по категориям основана на семантических, синтаксических, морфологических, словообразовательных, фонологических признаках⁵⁶. Поскольку омонимические слова принадлежат более чем к одной категории с несколькими значениями, в практике составления МА были разработаны разные форматы выделения основных морфем:

- 1) разделение слов на классические части речи при выделении основного типа;
- 2) отличие основного типа на основе структурно-модельного морфологического типа.

В то время как классификация первого метода применяется к процессу работы филолога, второй метод применяется к процессу программирования⁵⁷.

Все основные морфемы делятся на пять морфологических групп на основе структурно-модельной морфологической классификации аффиксальных морфем, таких как N, A, V, D, S, A – это группа морфем, которая присоединяет суффиксы, связанные с качеством, справа. V – это группа основ, которая принимает суффиксы, прикрепляющие к глаголу справа. D – группа из N суффиксов справа и группа суффиксов, образующих числовую форму. S – это класс морфем, которые могут присоединять модальные аффиксы только справа. Перечисленные морфологические группы далее подразделяются на подгруппы в соответствии с их способностью присоединять морфемы внутри себя⁵⁸.

В тюркских языках, в том числе и в узбекском, слово имеет разное значение, в зависимости от его положения в контексте. Непозиционная морфема – это морфема, которая принадлежит семантической группе, независимо от ее положения в предложении⁵⁹. Основа морфемы принадлежит к группе слов, обозначающих значение объекта, независимо от его окружения. Позиционная морфема принадлежит к разной семантической группе в зависимости от ее положения в предложении. Такие морфемы могут включать в себя имена существительные, которые представляют вещество-минерал: *золото, дерево, железо*. Сравните: *Уста ёғочни арралади//Богбон гулни ёғоч тувакка ўтқазди*⁶⁰. Позиционные основы включают как слово, обозначающее знак объекта, так и слово, обозначающее знак действия.

Лингвистическое обеспечение МА требует наличия словаря основ (лексикон). Узбекский язык, основанный на «Орфографический словарь узбекского языка»⁶¹ или пятитомнике «Толковый словарь узбекского

⁵⁶ Тот же источник. – С. 75.

⁵⁷ Тот же источник. – С. 76.

⁵⁸ Тот же источник. – С. 76.

⁵⁹ Тот же источник. – С. 79.

⁶⁰ Тот же источник. – С. 80.

⁶¹ Ўзбек тилининг имло лугати. (Тузувчилар: Э.А.Бегматов, А.П.Мадвалиев; Н.Махмудов тахрири остида). – Тошкент: Akademnashr, 2013. – 520 б.

языка»⁶², содержащем 85 000 слов для лексики МА узбекского языка, на основе этих словарей словарная база может состоять из следующих: лексикон-имя существительное, лексикон глаголов, лексикон имя прилагательное, общий лексикон (число, местоимение, наречие, союз, послеслог, частица, междометие, модальных слов, подражание).

Раздел данной главы, озаглавленный «*Морфоанализатор узбекского языка единицы анализа, словоформы и морфемы*», описывает различие между грамматической категорией (ГК), грамматическим значением (ГЗ), граммемой, квазиграммемой, дериватомой в узбекском МА.

Грамматическая категория (ГК) – это система взаимно противоречивых форм, объединённых на основе определенной общности значений; это не простая арифметическая сумма форм, а новое целое, состоящее из устойчивой относительной определенного типа формы⁶³. Следовательно, набор форм, представляющих один и тот же ГЗ, составляет грамматическую категорию. Первым шагом в процессе морфоанализа является определение того, к какой ГК принадлежит грамматическая форма, поэтому в МА сначала необходимо различать свойственные для языка ГК.

В узбекском языке разделяют 11 видов ГК⁶⁴: притяжательная форма, падеж, число, залог, время, наклонение, лицо-число, степень, положительная-отрицательная форма, форма действия, спряжение глагола. Основываясь на приведенных выше 11 грамматических категориях в обозначении ГЗ в МА узбекского языка, процесс автоматического морфоанализа дает точные результаты.

В узбекском языке ГЗ представляется следующим методом:

- 1) добавления к основе ГЗ суффикса: *китобни*;
- 2) с помощью вспомогательных слов: *Биз келажакка ишонч билан қараймиз*;
- 3) с помощью интонации: *Бу китоб (қайси?)*. – *Бу – китоб (нима?)*;
- 4) повтор слова.

МА умеет анализировать первый и второй способы выражения ГЗ. Для этого требуется информация о грамматической форме и вспомогательных словах в базе данных. По доступности ГК делится на три группы:

1. Природа слова – это уже существующее ГК. Слово означает ГЗ, если оно принадлежит к категории. Следовательно, первая граммема, прикрепленная к каждому слову, происходит от ГЗ, которая уже существует по своей природе⁶⁵.

2. ГЗ создаётся с помощью грамматической формы. Такое значение существует в самостоятельном слове и образуется как грамматическим суффиксом, так и вспомогательным словом⁶⁶. Отсюда следует, что в

⁶² Ўзбек тилининг изоҳли луғати. 5 жилд. – Тошкент: “Ўзбекистон миллий энциклопедияси” Давлат илмий нашриёти, 2006-2008. – 680 б.

⁶³ Замонавий ўзбек тили. Морфология. (Масъул муҳаррир: Р.Р.Сайфуллаева). – Тошкент: Мумтоз сўз, 2008. – 522 б. – Б. 86.

⁶⁴ Замонавий ўзбек тили. Морфология. (Масъул муҳаррир: Р.Р.Сайфуллаева). – Тошкент: Мумтоз сўз, 2008. – 522 б. – Б.186.

⁶⁵ Тот же источник.

⁶⁶ Тот же источник.

морфоанализе вторая граммема определяется на основе грамматических средств, добавленных к слову.

3. Грамматическое значение, которое проявляется во фразе и структуре предложения. Самостоятельное слово – это главное или второстепенные члены предложений, слово-обращение, вступительное слово, выделенные члены предложения или же является словом в предложении. Вспомогательное слово, с другой стороны, связывает слово или предложение, придавая ему дополнительное значение⁶⁷. Это значение формируется только в структуре предложения, в слове вне предложения такого ГЗ нет. Такое ГЗ не анализируется в процессе морфоанализа, но это означает, что следующим этапом анализа является задача синтаксического анализа. Следовательно, такое ГЗ не определяется как граммема.

По своему отношению к грамматической категории, то есть с точки зрения создания или не создания парадигмы, словоформы делятся на два типа: категориальные и некатегориальные. Категориальные формы – это формы, которые объединяются в систему грамматических форм, характерных для определенной группы слов, и, следовательно, образуют грамматическую категорию: в именах существительных, грамматические категории таких как: формы числа, принадлежности и падежа, время в глаголах, личностно-числовые формы. В морфоанализаторе они обозначаются как граммема. И.А.Мельчук сторонник за употребление термина грамматическое значение в отношении как реляционной, так и некатегориальной форм, а также словообразовательных форм. По этой причине ученый считает, что грамматическое значение состоит из синтаксических форм и словообразователей: первое называется граммемой, второе дериватомой. Отличительная особенность этих двух концепций заключается в следующем⁶⁸:

Дериватемы имеют тенденцию	Грагмемы имеют тенденцию
1) быть конкретными; 2) иметь относительно ограниченную сочетаемость; 3) иметь нестандартные средства выражения; 4) оставаться вне сферы действия синтаксических правил; 5) допускать амальгамирование с лексическими значениями; 6) выражаться показателями, располагающимися ближе к корню, чем показатели грамем; 7) изменять часть речи исходной основы.	1) быть абстрактными; 2) иметь относительно широкую сочетаемость; 3) иметь стандартные средства выражения; 4) фигурировать в синтаксических правилах; 5) избегать амальгамирования с лексическими значениями; 6) выражаться показателями, располагающимися дальше от корня, чем показатели дериватомы; 7) не изменять часть речи исходной основы.

Некатегориальные формы тоже имеют определенное грамматическое значение, но они не сливаются в единную систему, они обладают

⁶⁷ Тот же источник.

⁶⁸ Мельчук И.А. Курс общей морфологии. Т.1, Москва – Вена: «Языки русской культуры», Венский славистический альманах, Издательская группа «Прогресс», 1997. – С.240 – 287.

определенной степенью самостоятельности. К некатегориальным формам относятся уменьшительные и ласкательные формы имена существительных, формы избытка/недостатка в прилагательных. Такие формы определяются в морфоанализаторе как квазиграммама.

И.А.Мельчук описывает квазиграммему как важное промежуточное явление. Разграничение между граммемы и дериватемы не абсолютно. Существуют значения, выражаемые регулярно, но не обязательные; не входя ни в какую словоизменительную категорию, они должны были бы считаться дериватами, если бы не их неперемнная регулярность. Во всех остальных отношениях они больше похожи на граммы: они абстрактны, обнаруживают неограниченную или, по крайней мере, достаточно широкую сочетаемость, имеют стандартные средства выражения, фигурируют в синтаксических правилах и никогда (или почти никогда) не амальгамируются с лексическими значениями; среди них есть и такие, что их показатели располагаются дальше от корня, чем показатели некоторых граммем⁶⁹. Кроме того, что еще важнее, они не подвергаются фразеологизации, т.е. они всегда строго композиционны. Предлагаем называть грамматические значения проиллюстрированного типа квазиграммемами. В квазиграммы могут быть включены все грамматические формы, не входящие в грамматическую категорию. Лексические формы присущие имени существительного: форма принадлежности *-ники*, аффиксы места *-даги*, сопредельная форма *-гача*, уменьшительно-ласкательная форма; уровень степени прилагательных *-роқ*, формы действия, которые означают мощность/немощность.

В связи с национальным характером тюркского языка, грамматические формы узбекского языка изучаются специалистами в трёх видах. (1) Рассматривается как квазиграммама, категория степени, среди формаобразователей, категория отношений граммема, образующие типы значений числа. (2) Все словоизменительные морфемы определяются как граммы. (3) Лексико-синтаксическое формообразование влияет как на лексическое значение лексемы, так и определяет ее синтаксическую функцию. Причастие, наречие, форма действия занимает в этой группе место по своей двойственной природе и имеет промежуточный третий статус⁷⁰. В МА эти формы также включаются в список граммем.

Раздел главы *«Правила морфотактики в морфоанализаторе узбекского языка»* обсуждает важность правил морфотактики, валентности и морфотак-тики в МА. Другой важный фактор в формировании МА – установление систем правил морфотактики. Морфотактика – это изучение свойства валентности морфем в процессе словообразования или синтаксической связи, их морфемного отношения⁷¹. Морфотактика – неотъемлемая часть раздела словообразования, изучающего связность

⁶⁹ Тот же источник.

⁷⁰ Тот же источник.

⁷¹ <http://www.slovorod.ru/russian-wordform.html>

морфем, валентные свойства и их взаимосвязь между морфемами⁷². Аффиксальные морфемы имеют собственную среду окружения, добавляются только к словам или основам в определенном диапазоне, образуя ту или иную словоформу. Такой особый процесс словообразования и формообразования называется морфемным отношением⁷³. Поскольку валентность – это хорошо изученный вопрос⁷⁴ в узбекской лингвистике, мы не будем останавливаться на данном вопросе.

Оптимальный способ создания морфотактических правил МА – разработка комбинационной модели словообразовательных морфем, включение их в языковую базу. На самом деле, количество словообразовательных морфем внушительное. Расположение словообразовательных аффиксов также является важным процессом в формировании морфотактических правил, моделирование локализации словообразователей – следующий этап после морфотактики. На наш взгляд, объем работы по морфологическому анализу увеличивается, если не уделять особого внимания фонетическим явлениям, которые происходят в основе (корень) при добавлении морфемы к слову. Наблюдения за методами определения морфотактических правил в программах анализа морфем⁷⁵ показали, что все аффиксальные морфемы, доступные в языке, помещаются в базу данных программы, после чего активируется функция определения морфотактических правил. Конкретный суффикс определяется из списка аффиксальных морфем. После этого добавления морфему можно ставить, нажимая кнопку сохранения. В этом порядке морфемы, которые могут быть смежными справа, присоединяются к каждой из дополнительных морфем. На основании наших наблюдений мы обнаружили, что этот метод является наиболее удобным способом установления морфотактических правил. В компьютерной лингвистике различают регулярные и нерегулярные морфемы⁷⁶. Некоторые суффиксы могут быть добавлены ко всем корням, принадлежащим к одной морфологической группе. Например, суффикс *-ga* добавляется к корню в морфологической группе N. Такие суффиксы называются регулярными морфемами. Морфемы, которые не могут быть добавлены ко всем основам, принадлежащим к одной морфологической группе, включают возвратные и страдательные относительные суффиксы (*-ин, -н, -ил, -л*). Они потенциально добавляются к морфеме глагольной основы, но на практике некоторые слова не могут принимать текущий суффикс.

⁷² http://window.edu.ru/catalog/pdf2txt/516/25516/8411?p_page=10

⁷³ http://window.edu.ru/catalog/pdf2txt/516/25516/8411?p_page=10

⁷⁴ Абузалова М.К. Субстанциал морфология, валентлик ва синтактик курилма: Филол. фан. доктори (DSc) диссертацияси. – Самарқанд, 2018. – 224 б.; Мухамедова С.Х. Ўзбек тилида ҳаракат феълларининг семантик ва валентлик хусусиятлари. Докт. дисс. – Тошкент, 2007.; Менглиев Б. Морфологик воситаларнинг маъновий хусусиятлари ва семантик имкониятлари. Филол. фан. номз. дисс. автореф. – Тошкент, 1996; Шахобиддинова Ш. Грамматик маъно талқини хусусида. Номз. дисс. – Самарқанд, 1993.

⁷⁵ <http://tatmorph.pythonywhere.com/> (25.11.2020)

⁷⁶ Формальные модели и программные инструменты компьютерной обработки татарского языка. / Р.Р.Гатауллин, А.Р.Гатауллиев, О.А.Неврозова, Д.Р.Мухамедшин, Д.Ш.Сулейманов, Б.Э.Хакимов, А.Ф.Хусаинов. – Академия наук РТ, Институт прикладной семиотики АН РТ. – Казань: Изд-во Академии наук РТ, 2019. – 260 с. – С.81-83.

Например, корень *безамоқ* принимает оба суффикса залога: *безанмоқ*, *безалмоқ*. Нет особой закономерности добавления/недобавления этих дополнений к корню. Определение возможности добавления/недобавления дополнений залога к корню упрощает процесс морфоанализа, ускоряет синтез. Это означает, что в узбекском языке есть граммы, которые добавляются к определенным словам, которые должны быть отсортированы отдельно среди суффиксов выбора слова. В противном случае процесс морфологического анализа не даст правильного результата.

Глава IV диссертации озаглавлена «**Проблемы лингвистического моделирования морфемной среды в словах для морфоанализатора узбекского языка**». В главе «*Лингвистическое моделирование формообразовательные, словоизменительные и среды частиц в форме аффиксов*» анализируются проблемы моделирования среды ФОМ, СОМ и морфем в форме частиц. Порядок грамматического устройства в структуре слова имеет определенную закономерность⁷⁶. Поэтому в частях речи среда ФОМ требует специального моделирования.

Глагольная форма ФОМ *-(u)ш* выполняет функцию действия и взаимно-совместного залога. Чтобы различать такую грамматическую омонимию, в базу данных должны быть включены следующие позиционные модели:

1) в позиции $V+f_1v$ или $V+f_1v+f_2v$ форма *-(u)ш* выполняет функцию действия: *чиқ+ар+иш*; *том+из+иш*, *кий+гиз+иш*, *кий+дир+иш*;

2) в позиции $V+f_1v+p_1v$ ёки $V+f_1v+f_2v+p_1v$ форма *-(u)ш* выполняет функцию взаимно-совместного залога: *кел+иш+ди*, *суришитир+иш+ди*, *ўтир+иш+ди*.

Глагол+взаимно-совместный залог+взаимно-совместный залог: *сўзла+ш+иш*. В данной позиции *-ш* имеет форму взаимно-совместного залога.

3) имеет значение залога в словоформах *жойлашди*, *керишди*, *қоришди*, что связано с переходными/непереходными проблемами глаголов (конечно, эта проблема теоретически ожидает своего решения)⁷⁷; поэтому эту ситуацию нужно определять не на основе моделирования, а на основе правила исключений.

В зависимости от положения относительной формы деепричастия *-(u)б* образует омонимию с формой прошедшего времени. Анализируем позиции:

1) суффикс деепричастия в форме *ўқиб*, *бориб* выполняет функцию вспомогательного глагола в форме *ўқиб кўрмоқ*, *гуллаб қўймоқ* при добавлении к слову связывает главный также вспомогательный глагол, а в сложных глаголах форма *олиб бормоқ*, *кириб чиқмоқ*, связывает главный также вспомогательный глагол и выполняет функцию деепричастия. Значит, $V+f_1v(u)б$, то есть, {глагол-основа + *(u)б* + 0} то, *-(u)б* = глагол в форме деепричастия; если $V_1 + f_1v(u)б + V_2 + f_1v + f_1v/p_1v$, то есть {глагол-основа + *(u)б* + глагол-основа + ФОМ/СОМ} то, *-(u)б* = форма деепричастия связывающая главный и вспомогательный глагол;

⁷⁶ Mengliyev B., Xoliyorov O'., Abdurahmonova N. O'zbek tilidan universal qo'llanma. (Qayta ishlangan 3-nashri). – Toshkent: Akademi nashr, 2014. – 389 b. – B.134.

⁷⁷ Тот же источник.

2) в словах *ўтирибман, борибсан, узоқлашибмиз* выполняет функцию аффикса прошедшего времени. Моделируем позицию следующим образом: если $V+f_1v(u)б+p_1v$, то есть {глагол-основа+(и)б+СФО} то, $-(и)б$ = форма прошедшего времени. В словах *шолча, богча (болалар богчаси), кўрпача* появление нового лексического значения *-ча* ФОМ не считается, слова не разделяются на значимые части⁷⁸. Такие слова вносятся в словарь исключения; *-ча* не подлежит морфемному анализу. При анализе справа налево *-ча* находится в качестве аффикса, но при проверке корневой части, в словаре исключения данные слова находятся в качестве основы; анализ завершается определением того, что текущие слова приравниваются одной лемме.

Суффикс *-лар* при обозначении уважительного смысла после формы принадлежности располагается перед формой принадлежности множественности. Сравните: *опамлар//опаларим*. Также в словоформе *опаларим -лар* обозначает обобщённость по отношению к личности, а в словоформе конкретизм. В морфоанализе отличие этого состояния обеспечивается следующей моделью. Когда $f = -лар$: 1) если $\{R+f+r\}$ то, f = множество; 2) если $\{R+r+f\}$ то, f = уважительный смысл.

В современном узбекском литературном языке соблюдается следующий порядок морфем личности⁷⁹:

- 1) глагол-основа + {-ган /-(и)б / -яп / -а / -й /-моқчи} + {-ман /-сан / -0 / -миз / -сиз/ -(ди)лар}.
- 2) глагол-основа + {-ди /-са} + {-м /-нг / -0 / -к / -нгиз/ -(ди)лар}.
- 3) глагол-основа + {-ай/-0 / -син / -айлик / -гин/ -(и)нгиз / -(син)лар / +-ишсин}.

Для морфологического анализа мы моделируем позицию данных суффиксов следующим образом:

- 1) $V+f_1v+r_1f$. здесь: $f_1v = \{-ган /-(и)б / -яп / -а / -й /-моқчи\}$; $r_1f = \{-ман /-сан / -0 / -миз / -сиз/ -(ди)лар\}$;
- 2) $V+f_2v+r_2f$. здесь: $f_2v = \{-ди /-са\}$; $r_2f = \{-м /-нг / -0 / -к / -нгиз/ -(ди)лар\}$;
- 3) $V+ r_3f$. здесь: $r_3f = \{-ай/-0 / -син / -айлик / -гин/ -(и)нгиз / -(син)лар / +-ишсин\}$.

Видно, что словоизменительные (синтаксические формообразующие) морфемы, характерные для существительных и глаголов, образуют грамматическую омонимию. Поскольку синтаксический формообразователь всегда стоит в конце слова, некоторые из них остаются в поиске справа налево в МА, то есть анализатор находит форму *-нгиз* как синтаксический формирователь в словах *бордим* и *укам*, *келсанг* и *уйқунг* на основе информации в базе данных. Именно здесь возникает вопрос: является ли данная синтаксическая форма характерной для существительного или глагола? МА запускает второй алгоритм, чтобы найти ответ, для которого

⁷⁸ Тот же источник. – Б. 175.

⁷⁹ Замоновий ўзбек тили. Морфология. (Масъул муҳаррир: Р.Р.Сайфуллаева). – Тошкент: Мумтоз сўз, 2008. – 522 б. – Б. 420-426.

требуется модель. Форма синтаксического формообразователя определяется следующей моделью:

- 1) если $N/A/Num/Pr + r_1n$, то есть, {имя-основа + $-(u)m/-инг/-(u)миз$ } то, r_n = форма принадлежности;
- 2) если $V + r_f$, то есть, {глагол-основа + $-(u)m/-инг/-(u)миз$ } то, r_f = указатель личности-числа.

Вторая часть главы называется «Проблемы моделирования словообразовательных морфем в словесной среде». Хотя структура слов в тюркском языке сложна, она состоит из основы и суффикса. Данной проблеме посвящены многие научные работы: словообразование, некоторые вопросы морфологии словарной базы (Н.Д.Арутюнов)⁸⁰, морфологическая структура слов (Н.А.Баскаков⁸¹, Е.В.Севортян⁸²), морфологическая структура слов в турецком языке (М.С.Джикия⁸³), природа тюркской агглютинации в трудах А.Н.Кононова⁸⁴. Для того чтобы морфоанализатор мог идентифицировать производное слово, производную часть и производную основу, необходима точная информация о структуре производного слова в его лингвистическом обеспечении. В структуре производного слова есть два разных компонента (два разных явления), которые отличаются друг от друга значением и функцией, то есть в производном слове: 1) основа словообразования (или мотивационная основа); 2) словообразовательный формант. Например, в словах *туртки, ишчан, қорай турт, иш, қора* – это основа словообразования (также известная как словообразовательная основа)⁸⁴.

В словаре основ анализатора все основы на узбекском языке хранятся с информацией о принадлежности к частям речи, то есть все основы являются леммой, а с помощью прикрепленного тега получаем информацию о том, к какой группе слов они принадлежат.

МА определяет и анализирует морфемы справа налево. На этом этапе МА должен определить три вещи: основу слова, производную основу и словообразовательный аффикс. Их обнаружение происходит по следующему алгоритму:

- 1) корень присутствует в словаре баз, значит, основа найдена;
- 2) на основании информации о суффиксах в базе данных анализ ведётся справа налево: идентифицируется словообразовательный суффикс;

⁸⁰ Арутюнов Н.Д. Некоторые вопросы образования и морфологии основ слов. – Филологические науки, 1958. – № 1.

⁸¹ Баскаков Н.А. Морфологическая структура слова и части речи в тюркских языках. – Советское востоковедение, 1957. – № 1.; Баскаков Н.А. Порядок и иерархия аффиксов в основе слова в тюркских языках. Баскаков Н.А. В сб.: Морфологическая типология и проблема классификации языков. М.-Л., 1965.

⁸² Севортян Э.В. Морфологическое строение слова в связи с другими его характеристиками. – В. Кн.: Тюркологический сборник. – М., 1972.

⁸³ Джикия М.С. Морфологическая структура слова в турецком языке: автореф. дисс. ... канд. филол. наук. – Тбилиси, 1975.

⁸⁴ Кононов А.Н. О природе тюркской агглютинации. – Вопросы языкознания, 1976. – № 4.

⁸⁵ Тот же источник.

3) следует заметить, что часть перед образующим основанием является основанием образования: основа образования находится⁸⁵.

Определим основную инвариантную модель словообразования следующим образом: $R+d$. В этом случае R – любая основа; d – это дериватема, т.е. словообразующая морфема. Базовая инвариантная модель $R + d$ встречается в следующей субмодели: R_v+d ; R_n+d ; R_a+d ; $R_{num}+d$; $R_{pr}+d$; $R_{adv}+d$; $R_{sim}+d$.

Производные слова, такие как *терим, гавдали, гулла, кийимлик, бирлашма, теримчи, теримчилик; асаларичи, белбогли; беш йиллик, темирйўлчи, ўзибўларчилик* выглядят следующим образом. Морфоанализатор должен определять словообразующие морфемы при анализе справа налево: *-им, -ли, -ла, -лик, -ма, -чи, -чилик, -арчилик*. При распознавании аффиксов с одним слогом проблем не будет, но начальную форму *-лик* можно различить в правом чтении форм *-чилик, -арчилик*. Итак, мы добавляем правило и модель в базу данных. Если {корень + *лик*} то, *-лик* может стать словообразовательным аффиксом. Модели: 1) если { $R + d_1$ (*лик*)} то, *лик* = d ; 2) если { $R + d_1 + d_2$ (*лик*)} то, $d_1 + \text{лик} = d$.

Инвариант образования имени существительного от имени существительного имеет модель R_n+d_n , инвариант образования имени прилагательного от имени существительного имеет модель R_n+d_a , инвариант образования наречия от имени существительного имеет модель R_n+d_{adv} , инвариант образования глагола от имени существительного имеет модель R_n+d_v , инвариант образования от имени прилагательного-основы имеет модель R_a+d , инвариант образования от имени числительного-основы имеет модель $R_{num}+d$, инвариант образования от местоимения-основы имеет модель $R_{pr} + d$, инвариант образования от наречия-основы имеет модель $R_{adv}+d$, инвариант образования от подражания-основы имеет модель $R_{sim}+d$, инвариант образования от глагола-основы имеет модель R_v+d . Следует отметить, что при словообразовательном анализе МА полный анализ словообразования не может быть осуществлен путем выявления только морфологических форм, то есть путем определения образовательной части. Типы словообразования также должны быть проанализированы для проведения качественного анализа. Исходя из этого, в исследовании было разработано 38 субмоделей инвариантных моделей, поскольку грамматический анализ в МА выполняется на основе субмоделей.

$R+d_1+d_2$ и окружение словообразовательных морфем в другой сложной субмодели. Как было сказано выше, в узбекском языке к основе можно добавить более одной словообразовательной морфемы: у нее есть взаимный порядок и цепочка словообразовательных морфем. Основной инвариант $R+d_1+d_2$ отражается с помощью следующих субмоделей: $R_n+d_1+d_2$; $R_a+d_1+d_2$; $R_{num}+d_1+d_2$; $R_{pr}+d_1+d_2$; $R_{adv}+d_1+d_2$; $R_{sim}+d_1+d_2$; $R_v+d_1+d_2$. Здесь, глагол, имя существительное, имя прилагательное, числительное, местоимение, наречие,

⁸⁶ Степанова М.Д. Вопросы моделирования в словообразовании и условия реализации моделей. – Вопросы языкознания, 1975. – № 4.

подражание образовано добавлением к основе двух словообразовательных морфем. Модели были разработаны с точки зрения распознавания формы МА, идентификации базового набора, порядка морфем, добавляемых после основы, и анализа влияния этого порядка на значение.

В разделе «О некоторых проблемах моделирования среды словообразовательных аффиксов» исключения в словообразовательных моделях изучаются при анализе вопросов лингвистического моделирования омонимии, многозначных словообразовательных морфем.

Спорные вопросы словообразования в узбекском языке были специально проанализированы⁸⁷, и эти работы были использованы при формировании моделей МА. А.Ходжиев пишет, что расположение образователя при изучении словообразования может быть незначительным при изучении феномена, но в морфоанализе этот процесс важен. Если словообразователь расположен после основы, т.е. справа, анализатор определит словообразовательный аффикс при анализе справа налево. Например: *супурги, кийим, сузма, тугун, қиринди, ювинди, қисқич, сузгич, илмоқ, қуймоқ, ўсимта, босқон, қирмоч, овунчоқ, кекирдак, қовурдоқ, ўроқ, суюнчи, ёпинчоқ*. Данные производные слова анализируются справа налево, определяя основу построения и образующуюся часть в форме на основе информации в базе данных и отображает результат *супур+ги, кий+им, суз+ма, туг+ун, қир+инди, юв+инди, қис+қич, суз+гич, ил+моқ, қуй+моқ, ўс+имта, бос+қон, қир+моч, овун+чоқ, кекир+дак, қовур+доқ, ўр+оқ, суюн+чи, ёпин+чоқ*. Однако в случае отсутствия шаблона {словообразователь+основа образования} в базе данных анализатора словообразователи которые расположены до основы образования при анализе префиксов не получают нужный результат. Следовательно, если список аффиксоидов и аффиксов находится в базе данных, МА может выполнять анализ на основе моделей.

Но есть ряд суффиксов, омонимия которых, определяется с помощью модели конкретного слово/словоформы. Например: в словах *қорай* и *ўқий* омонимный суффикс *-й* является глаголообразующим, также характерно для глагола является синтаксическим формообразователем. Мы знаем, что глаголообразующий суффикс создает глагол из основы, которая не является неглагольной основой. Мы моделируем на основе данного правила:

- 1) если {неглагольная основа + *-й*} то есть, *й* = образователь глагола;
- 2) если {глагол-основа + *-й*} то есть, *й* = образователь лексической формы (формаобразователь).

Итак, путь фильтрации на основе модели был найден. Морфема *-кин* выполняет функцию прилагательного в слове *кескин* и лексической формы глагола в слове *тўккин*. В обоих случаях к основе (к корню) добавляется глагол, который не может быть использован в модели по его положению относительно основы.

⁸⁷ А.Ходжиев, А.Нурманов, Н.Махмудов. Спорные вопросы в системе словообразования тюркских языков. // Советская тюркология, 1989. – №2.

При моделировании среды и порядка морфем в слове можно сделать следующие выводы:

- 1) существует определенный порядок морфем; обычный порядок в них следующий: основа + словообразующая морфема + ФО морфема + СФ морфема;
- 2) среда словообразовательных морфем связана с агглютинативным характером узбекского языка, и имеется много случаев их последовательного сложения, внутреннее и внешнее словообразование также различно;
- 3) в узбекском языке не менее 4 слов, составляющих слово, что связано с несоблюдением принципа лингвистической экономики;
- 4) все словообразователи располагаются после основы, перед ФОМ;
- 5) если в составе слова есть два словообразователя, сначала добавляется конкретное СОМ имя существительное, затем добавляется абстрактное СОМ имя существительное;
- 6) в узбекском языке производные словообразователи добавляются после словообразователей;
- 7) слово со слабым смысловым характером занимает позицию после слова;
- 8) вначале идёт непродуктивный словообразователь, затем располагается продуктивный словообразователь;
- 9) если слово содержит уменьшительно-ласкательный суффикс, вначале добавляется уменьшительная форма, затем ласкательная форма;
- 10) два функционально одинаковых ФОМ добавляются последовательно, т.е. приобретают плеонастический характер;
- 11) порядок лексических форм глагола следующий: {относительная форма + отрицательная форма + функциональные формы глагола + временная форма};
- 12) СФО считаются истинными словообразователями, потому что они образуют синтаксическую функцию слова;
- 13) в узбекском языке предлоги всегда добавляются в конец слова (есть случаи, когда они добавляются последовательно);
- 14) нарушение обычного порядка в структуре слов связано с определенными языковыми факторами.

ЗАКЛЮЧЕНИЕ

1. Лингвистическое программное обеспечение – это совокупность языков, используемых в программном обеспечении информационных технологий, метод, средства, применение и использование их создания. Набор инструментов, используемых при автоматической обработке текстов на естественном языке, языковой процессор. Лингвистическое обеспечение информационных систем – это набор инструментов, структура объекта информации, действия, выполняемые над понятием, которое используется для его функционирования, выраженные на формальном, естественном языке.

2. Лингвистическое программное обеспечение – это набор программного обеспечения и данных для аудиоданных, изображений, анализа текста, обработки, хранения на естественном языке. Он разделен на электронный словарь, орфографический корректор, поисковую систему, систему машинного перевода, программное обеспечение для управления памятью переводчика, систему анализа/синтеза речи, систему голосового перевода. Лингвистическое обеспечение информационно-поисковой системы включает алфавит, микросинтаксис, словарь, тезаурус, классификатор; инструменты для адаптации разных языков; практические инструменты.

3. Задача морфологического анализа – автоматически определить категориальный признак слова в тексте, определить его принадлежность к лексико-грамматическому классу. Синтаксический анализ текстовой информации позволяет выделить смысловой элемент предложения (принадлежность к группе, предикативная основа). Лексический анализ выявляет формы речи, лексического выражения (ассимиляция, сленговое слово), которые указывают на специфику языка. В процессе автоматического анализа компьютер создает лексико-морфологический, синтаксический, смысловой образ текста на своем языке. Практическая значимость базы данных, программ анализа определяется повышением качества ИПС и машинного перевода, лингвистической разметки единиц языкового корпуса, а также его работой в качестве независимого анализатора. Морфологический анализатор - это словоформа в словаре, предназначенная для сравнения отдельных слов, средство выражения грамматической особенности слова.

4. Лингвистическое обеспечение морфоанализатора хранится в виде базы данных: база данных состоит из словаря основ и словаря правил. База данных – это большой набор данных, в котором данные сортируются и хранятся в табличной форме в системе баз данных. Словарь основ состоит из лексики языка, грамматического словаря, составляющего список грамматических форм в языке; базы аффиксов; правила классификации; Правила совместимости алломорфов находятся в базе данных. Лингвистическое обеспечение морфоанализатора в случаях словоизменения в узбекском языке (отражение фонетических изменений на границе корня и аффикса); включает информацию о дополнительных, инвариантных

лексемах, а также о традиционных синтаксических формах; покрывает основу омонимических лексем.

5. К системе морфоанализатора предъявлены следующие требования: определить части речи и другие морфологические знаки, различить корень/основу и аффиксы, понять текст, анализировать слова-исключения, предоставить результат в виде файла. Основные единицы морфоанализатора – морф, алломорф, слово. Морф – это минимальное явление, которое имеет ровно одну морфологическую особенность, форма которой случайным образом подобна; алломорф – сумма морфов морфемы с разным фонетическим содержанием; лексикон – это последовательность алломорфов, в которой существует морфотактически определенная связь между ее компонентами. С другой стороны, морфотактические правила служат для определения отношений между базовыми и аффиксальными группами.

6. Морфологический анализ состоит из основных блоков, таких как стемматизация, лемматизация, грамматизация. Стемматизация – это процесс поиска основы искомого слова. Лемматизация – это процесс преобразования слова в лемму, что важно для флективных языков, потому что в агглютинативных языках «основа» слова равна лексеме. Только у флективной словоформы наблюдается несовпадение леммы и стема. Ещё одним фактором повышения качества анализа является идентификация позиционных и непозиционных морфем. Позиционная морфема принадлежит семантической группе независимо от ее положения в предложении. Непозиционная морфема принадлежит к разной семантической группе в зависимости от ее положения в предложении. Формирование списка позиционных морфем в базе данных, моделирование их локализации повышает качество морфологического анализа.

7. По отношению к грамматической категории словоформы делятся на два типа: категориальные и некатегориальные формы. Категориальная форма – это форма, которая объединяется в систему грамматических форм, специфичных для определенной группы слов, образуя грамматическую категорию, которая определяется как граммема. Некатегориальные формы также имеют определенное грамматическое значение, но не сливаются в одну единую систему, обладают определенной степенью самостоятельности, их называют квазиграммема. Словообразовательная форма в МА – дериватема. Морфоанализ узбекского языка состоит из следующих этапов: (1) поиск возможного корня (основы); (2) найти возможную основу соответствия из словаря основ; (3) определить морфологический тип корня, найденного в базе данных; (4) анализировать правую часть (группы аффиксов).

8. Порядок грамматических инструментов в слове имеет определенную закономерность, а окружение в слове определяет его функцию. Для морфоанализа следует моделировать грамматические формы, которые имеют одинаковую форму, но разное окружение слов. Определение грамматической формы, форма которой такая же, как у слова, не может быть достигнуто даже с помощью морфоанализа, они должны подвергаться семантической/синтаксической автоматической обработке. Моделирование морфемного

окружения, образующего лексическую форму в слове, гарантирует, что в морфоанализе он отличается от словообразовательного, словоизменительного (синтаксического формообразователя). Морфоанализатор должен иметь информацию о структуре производного слова в языковом обеспечении, чтобы идентифицировать производное слово, образующую часть и основу образования. На основе моделей словообразования анализатор определяет три вещи справа налево: основу слова, основу образования и словообразовательный аффикс. Модель разработана с точки зрения морфоанализатора, распознающего форму, определения базовой категории, порядка морфем, добавляемых к основе и анализа влияния этого порядка на значение.

9. Отклонение от нормы литературного языка служит созданию индивидуальной лексики, характера главного героя. События, которые не типичны для литературного языка, но распространены в речи, также должны быть записаны в качестве эквивалентов в базе данных морфоанализатора, поскольку произведение искусства, устная речь (иногда диалектный) материал также может быть подвергнут морфологическому анализу с целью пользователя. Исключения в дополнительных морфемах необходимо включать в «словарь исключений». При составлении перечня таких форм наряду с литературной нормой необходимо учитывать ненормативные, но широко используемые формы.

10. 39 субмоделей, основанных на 9 инвариантных моделях, для морфоанализатора для определения среды морфем, образующих лексическую форму узбекского языка; на основе одной инвариантной модели для автоматического анализа словообразовательных суффиксов было разработано 7 представлений субмоделей и 60 представлений субмоделей. Для автоматического анализа сложного словообразования разработаны в основном инварианты и субмодели. В частности, чтобы различить добавление двух словообразующих морфем к основе, 7 субмоделей на основе 1 инварианта, 5 его представлений; была идентифицирована 1 инвариантная модель, чтобы дифференцировать добавление трех, четырех словообразовательных морфем к основе. Всего было разработано 13 инвариантов, 118 субмоделей и их представлений.

**SCIENTIFIC COUNCIL FOR AWARDED SCIENTIFIC DEGREES
DSc.03/30.12.2019.Fil.05.02 AT FERGANA STATE UNIVERSITY
TASHKENT STATE UNIVERSITY OF UZBEK LANGUAGE
AND LITERATURE**

KHAMROEVA SHAHLO MIRDJONOVNA

**LINGUISTIC SUPPORT OF UZBEK LANGUAGE'S MORPHOLOGICAL
ANALYZER**

10.00.11 – Theory of language, applied and computational linguistics

**ABSTRACT
OF THE DOCTORAL THESIS (DSc) ON PHILOLOGICAL SCIENCES**

Fergana - 2021

The theme of DSc dissertation is registered by the Supreme Attestation Commission under the Cabinet of Ministers of the Republic of Uzbekistan under № B2020.4.DSc/Fil227

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, English (resume)) has been placed on the website (www.fdu.uz) of the Scientific Council and on the website of "ZiyoNet" Information-educational portal (www.ziyo-net.uz).

Scientific consultant:

Mengliev Bakhtiyor Rajabovich
Doctor of Philology, professor

Official opponents:

Samsitova Lulza Khamzlovna
Doctor of Philology, professor

Iskandarova Sharifa Madalievna
Doctor of Philology, professor

Shaxabiddinova Shokhida Khashimovna
Doctor of Philology, professor

Leading organization:

Samarkand State University

The dissertation defence will be held on 28 » 06 2021 at 9⁰⁰ at the meeting of the Scientific Council DSc.03/30.12.2019.Fil.05.02 at Fergana State University (Address: 19, Murabbiylar street, Fergana, 150100. Phone: (99873) 244-44-02; Fax: (99873) 244-44-93; Web-site: www.fdu.uz; e-mail: fardu_info@mail.ru).

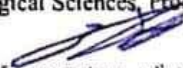
The dissertation is available at the Information Resource Center of Fergana State University (registered number № 112). (Address: 19, Murabbiylar street, Fergana, 150100. Phone: (99873) 244-44-02; Fax: (99873) 244-44-93; Web-site: www.fdu.uz; e-mail: fardu_info@mail.ru).

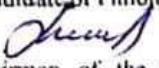
Dissertation abstract is distributed on « 17 » 06 2021.

(Mailing report № 112 on « 17 » 06 2021).




A.A. Kasimov
Chairman of the Scientific Council
awarding Scientific Degrees, Doctor of
Philological Sciences, Professor


V.T. Zakirov
Scientific secretary of the Scientific
Council awarding scientific degrees,
Candidate of Philological Sciences


A.B. Mamajonov
Chairman of the Scientific Seminar
under Scientific Council awarding
Scientific degrees, Doctor of
Philological sciences, Professor

INTRODUCTION (abstract of DSc thesis)

The purpose of the dissertation is to create a theoretical basis and models of linguistic support of the morphological analyzer of the Uzbek language.

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

The subject of research is the basics of modeling and systematization of linguistic supply units.

Task of the research consists of the following:

- to research the general characteristics of linguistic support tools and their role in the creation of morphological analyzers and language corpora;

- to study the tools of the process of automatic morphological analysis / synthesis of language units;

- to research the structure, composition and functions of linguistic support;

- to determine the general principles of morphological analyzer and its importance in the information retrieval system;

- to determine commonalities and differences of morphological analyzers of Turkic languages;

- identification of units, structure, composition and units of analysis of the Uzbek language morphoanalyzer;

- linguistic modeling of the lexical form-forming, syntactic-form-forming, word-forming morphemes, as well as the set of adverbs in the adverbial form.

Scientific novelty of the research consists of the following:

- specific functional aspects of linguistic analyzers such as orthocorrector, lemmatizator, morfoanalyzer, parser were determined by defining the structure and functions of the linguistic support of the information retrieval system in the Uzbek language;

- it was proved the morphological analyzer is the main means of linguistic annotation of the text and automatic processing of the natural language in machine translation;

- it was found that the functional capabilities of morphological analyzers developed in Turkic languages are different in terms of linguistic support, and common in terms of morphological analysis methods of automatic processing of language;

- combinatorial and positional linguistic support programs of morphotactics have been developed for words that differ from the general rules of word forms due to spelling changes in the Uzbek language;

- units, word-forming, lexical-forming, syntactic-forming morphemes and word-joining particles of additional-form loadings for automatic morphological analysis of Uzbek language are linguistically modeled for the software algorithm.

Scientific and practical significance of the research results. The research plays a theoretical role in the development of theoretical foundations for the creation of a morphological analyzer of the Uzbek language and its linguistic support, in the creation of research in the field of computational linguistics. The practical significance of the research is explained by the fact that in teaching

applied linguistics, computational linguistics, corpus linguistics can be used as a source in the development of programs, plans and descriptions of topics, as well as in the development of a morphological analyzer of the Uzbek language, a spellchecker, morphological markup of various types of corpus in the Uzbek language.

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 264 pages.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

LIST OF PUBLISHED WORKS

І бўлим (І часть; I part)

1. Хамроева Ш.М. Ўзбек тили морфоанализаторининг лингвистик таъминоти. Монография. – Германия: GlobeEdit, 2020. – 244 б.
2. Основы создание лингвистической базы морфологического анализатора узбекского языка. Монография. – Германия: GlobeEdit, 2020. – 273 с.
3. Khamroeva Sh.M. The author's lexicography and author's corpus approach // International Journal of Applied Research. – 2019, SP7. ISSN Print: 2394-7500 – ISSN Online: 2394-5869. – Impact Factor: 5.2 IJAR 2019; SP7: 26-29. ((35) CrossRef; (40) ResearchGate)
4. 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)
5. Khamroeva Sh.M. Morphotactic rules in the morphological analyzer of the uzbek language // “Middle European scientific bulletin”. – 2020. Volume 6. ISSN 2694-9970 – P. 45-49. ((23) Scientific Journal Impact Factor; (14) ResearchBib)
6. Хамроева Ш.М. Корпус лингвистикасининг бугунги ҳолати // Бухоро давлат университети илмий ахбороти. – Бухоро, 2019. – № 3. (10.00.00; № 1)
7. Хамроева Ш.М. Тил корпусларида семантик гуруҳ, тўда ва майдон таснифининг матнни семантик теглашдаги аҳамияти // Филология масалалари. – 2020. № 1. (1994–4233 ISSN). – Б. 81-94. – DOI: 10.36078/987654419 (10.00.00; № 18)
8. Хамроева Ш.М. Тил корпусини дастурлаш: сунъий тилни яратиш, тил синтаксиси ва семантикаси // Бухоро давлат университети илмий ахбороти. – 2020. № 1 (77). Б. 105-110. (10.00.00; № 1)
9. Хамроева Ш.М. Ўзбек тили корпусида морфологик разметка ва лингвистик модел (феъл ва от туркуми мисолида) // Бухоро давлат университети илмий ахбороти. – 2020. № 3 (79) – Б. 107-112. (10.00.00; № 1)
10. Хамроева Ш.М., Гулямова Ш.Қ. 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. Хамроева Ш.М. Муаллифлик корпусининг муштарак ва ўзига хос жиҳатлари // Сўз санъати халқаро журнали – 2020. № 2. – Б. 80-87. (10.00.00; № 31)
12. Хамроева Ш.М. Тил корпуси морфологик таҳлилида сўзшакллар разметкаси (айрим сўз туркумлари мисолида) // “Тил, таълим ва таржима” журнали. – 2020. № 4. – Б. 36-46. (10.00.00; №10)
13. Khamroeva Sh.M. The importance and specificity of written and oral, monolingual and multilingual corpora // Крымскотатарская филология: проблемы изучения и преподавания. Научный журнал. – 2020 г. № 1(4). – Б. 63-69.
14. Хамроева Ш. О построение морфологических анализаторов языковых корпусов узбекского языка // IV Междисциплинарная научная конференция “Конвенгертные технологии XXI: вариативность, комбинаторика, коммуникация». Симферополь, 2019 г ноябрь. – Симферополь, 2020. – 416 с. – С. 226-230.
15. Хамроева Ш. Использование морфоанализа языковых корпусов в преподавание морфологии // Реалізація компетентісно орієнтованого навчання в освіті: теоретичний і практичний аспекти: Зб. наукових праць за матеріалами Міжнародної науково-практичної конференції (м. Київ, 4 листопада 2019 р.), м. Київ / Ін-т педагогіки НАПН України. – Київ: Пед. думка, 2019. 422 с. – Б. 350-353.
16. Хамроева Ш. Особенности морфоанализа языковых корпусов // Страны. Языки. Культура: сборник материалов XI-й международной научно-практической конференции /

Под ред. проф. Абуевой Н.Н. Махачкала: ДГТУ, 2020. – 391 с. (ISBN 978-5-907249-44-8). С. 354-356.

17. Хамроева Ш. Некоторые вопросы морфотактических правил морфоанализатора узбекского языка // Инновационное развитие науки и образования. Международная научно-практическая конференция. – Казахстан. Павлодар. 2020 г, август. – Б. 94-97.

18. Хамроева Ш. Ўзбек тили морфологик анализаторини тузишда ассоциация ташки омонимликни фарқлаш механизми сифатида // “Ўзбек тилшунослигини ривожлантиришнинг долзарб муаммолари” Республика илмий-амалий конференцияси материаллари тўплами (2020 йил 21 август). – Тошкент: «O‘ZKITOBSAVDONASHRIYOTI», 2020. – Б. 95-101.

19. Хамроева Ш. Лингвистик таъминот: тадқиқот ва натижалар // O‘zbek tilshunosligining dolzarb masalalari: O‘zbek tiliga Davlat tili maqomi berilganligining 30 yilligiga bag‘ishlangan respublika ilmiy-amaliy konferensiya materiallari [Текст] 19-oktabr 2019-yil. / A.R.Axmedov. – Buxoro: “Sadriddin Salim Buxoriy” Durdona nashriyoti, 2019. - 136 с. – Б. 65-68.

20. Хамроева Ш. Инновацион технологияларни жорий этишда дастурий таъминотларнинг ўрни // “Бухоро вилоятини инновацион ривожлантириш муаммо ва ечимлар” мавзусида республика илмий-амалий анжумани материаллари. Бухоро, 2020. – 1220 б. – Б. 686-690.

21. Хамроева Ш. Морфологик таҳлил ахборотни қайта ишлашнинг илк босқичи сифатида // “Nutq madaniyati va o‘zbek tilshunosligining dolzarb muammolari” mavzusidagi xalqaro ilmiy-amaliy konferensiya materiallari. 2020-yil 4-may / [Mas‘ul muharrirlar: F.Usmonov, B.Ahmedov] – Andijon: Andijon davlat universiteti kichik bosmaxonasi, 2020. – 420. – Б.175-180.

22. Хамроева Ш. Корпус лингвистикасида разметкашнинг халқаро стандартлари // Замонавий таълимда рақамли тизимларни қўллаш: филология ва педагогика соҳасида замонавий тенденциялар ва ривожланиш халқаро илмий-амалий масофавий конференция. 1 май 2020 йил. – Тошкент: Tadqiqot, 2020. – 421 б. – Б. 46-50.

23. Хамроева Ш. Рус ва украин тиллари морфоанализатори: тадқиқ ва кузатувлар // «Глобальная наука и инновация 2020: Центральная азия» № 5(10). Август, 2020. – ISSN 2664-2271 – Серия Филологические науки». – Б.122-125.

24. Хамроева Ш. Регулярные и нерегулярные морфемы // “Современная филология: проблемы и перспективы”. Международной научно-практической конференции, посвященной Году башкирского языка, 85-летию со дня рождения д.ф.н., профессора, академика АН РБ Зиннура Газизовича Ураксина, 85-летию со дня рождения д.ф.н., профессора, академика АН РБ Марата Валиевича Зайнуллина. 18-20 ноября 2020 г. – Уфа, 2020. – 421 с. – С. 46-50.

II бўлим (II часть; II part)

25. Хамроева Ш. Ҳозирги ўзбек адабий тили фанидан таълим технологияси. “Ўрта махсус, касб-хунар таълими тизимида инновацион технологиялар” туркуми. – Тошкент: ReliablePrint, 2013. – 153 б.

26. Менглиев Б., Хамроева Ш., Тоирова Г., Жўраев Қ., Бахронова Д. Ўзбек тилининг миллий ва муаллифлик корпуси учун ЭҲМ дастури. – № DGU 05735 (муаллифлик гувоҳномаси). 23.10.2018.

27. Mengliyev B., Xamroyeva Sh. Korpus lingvistikasi: korpus tuzish va undan foydalanish (amaliy mashg‘ulot uchun qo‘llanma). – Globe Edit, 2020. – 57 b.

28. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Издатель: GlobeEdit (7 февраля 2020 г.). – 260 с.

29. Хамроева Ш. Сўзларнинг тузилишига кўра турлари. Содда, қўшма сўз ва уларнинг услубий қўлланилиши” мавзуси бўйича гуруҳларда ҳамкорликда ўқишнинг педогогик технологияси // Кичик гуруҳларда ҳамкорликда ишлаш педагогик технологиялар тўплами.

- “Ўрта махсус, касб-хунар таълими тизимида инновацион технологиялар” туркуми. 6-китоб. – Тошкент: Иқтисодиёт, 2013. – Б.33-36.
30. Ҳамроева Ш., Йўлдошева Д. “Мураккаб тузилишли сўзлар” мавзусини ўтиш // “Интерфаол таълим: муаммо ва ечимлар” мавзусидаги Республика илмий-амлаий анжумани материаллари тўплами I. – Бухоро, 2018. – Б.158-160.
31. Ҳамроева Ш., Менглиев Б. Тег как лингвистический инструмент // Материалы VI Международной научно-практической конференции «Global science and innovations 2019: Central Asia». VI ТОМ. – Нур-Султан, 2019. – С. 81-85.
32. Ҳамроева Ш. Интерфейс – корпус мундарижаси // Тилшунослик ва методика масалалари: “Яхшилар ёди ўчмайди сира”, доцент В.Эгамбердиевнинг 80 йиллигига бағишланган илмий-амалий конференция (2019, Бухоро). – Тошкент: Мухаррир нашриёти, 2019. – 216 б. – Б.70-72.
33. Ҳамроева Ш., Бозорова Л. Корпусни лойиҳалаш ва тузиш босқичининг технологик жараёни // Адабиётшунослик ва таржимашунослик нинг долзарб муаммолари: адабий жараён, қиёсий адабиётшунослик, услубшунослик ва тилшунослик масалалари мавзусидаги республика илмий-амалий конференция материаллари тўплами. Бухоро, 2019. – 492 б. – Б.195-198.
34. Ҳамроева Ш., Бозорова Л. Корпус менежери қидирув тизими сифатида // Умумфилологик тадқиқотлар: муаммо ва ечимлар. Республика илмий-амалий конференция материаллари тўплами. – Тошкент: Мумтоз сўз, 2019. – Б. 123-125.
35. Ҳамроева Ш., Менглиев Б. Тил корпусларини синтактик разметкашнинг лингвистик асослари // “Осиё мамлакатлари тамаддуни ва ипак йўли”. Халқаро илмий анжуман материаллари. – Самарқанд, 2019 йил 21 август. – 236 б. – Б. 36-39.
36. Ҳамроева Ш., Менглиев Б. Принципы создания интерфейса авторских корпусов узбекского языка (на примере авторского корпуса Абдуллы Каххара) // VII Международная конференция по компьютерной обработке тюркских языков “Turklang-2019”. – Симферополь, 2019 г. 3-5 октября. – 328 б. – Б. 36-39.
37. Ҳамроева Ш., Менглиев Б. Лингвистик таъминот воситаларининг умумий тавсифи ҳамда тил корпусини яратишдаги аҳамияти // Международная научно-практическая конференция «Развитие узбекского языка и проблемы тюркологии». – Ташкенте, 2019 г., 29 октября.
38. Khamroeva Sh., Mengliev B. The authorship lexicography and authorship corpus // Материалы международной научной конференции межнаучная интеграция: лингводидактический, лингвокультурологический и психолингвистический аспекты. 19-20 декабря 2019 год. – Sumqayit – 2019.– Б. 105-107.
39. Sharipova M., Xamroyeva Sh., Jo'rayeva B. O'zbek xalq maqollari korpusi interfeysi tavsifi // Материалы V Международной научно-практической конференции «Наука и образование в современном мире: вызовы XXI века». Секция 9. Филологические науки. II ТОМ. – Нур-султан, 2019. – 162 б. – Б. 21-23.
40. Abuzalova M., Xamroyeva Sh., Boltayeva N. Publitsistik metaforalar korpusi (PMK) interfeysi haqida // Материалы V Международной научно-практической конференции «Наука и образование в современном мире: вызовы XXI века». Секция 9. Филологические науки. II ТОМ. – Нур-султан, 2019. – 162 б. – Б. 17-21.
41. Xamroyeva Sh., Ubaydullayeva X. O'zbek tilida somatik iboralar korpusining interfeysi va manbalari // Материалы V Международной научно-практической конференции «Наука и образование в современном мире: вызовы XXI века». Секция 9. Филологические науки. II ТОМ. – Нур-султан, 2019. – 162 б. – Б. 6-8.
42. Ҳамроева Ш. Лингвистик таъминотнинг тузилиши, таркиби ва вазифалари // “Ўзбекистонда илмий-амалий тадқиқотлар” мавзусидаги республика 15-кўп тармоқли илмий масофавий онлайн конференция материаллари тўплами. 30 апрель 2020 йил. – Тошкент: Tadqiqot, 2020. – 289 б. – Б.149-151.
43. Khamroeva Sh., Mengliev B. Some comments on the semantic tagging of verbs in uzbek

- language corpus // Science and education scientific journal (ISSN 2181-0842) “Илм-фан ва таълимнинг ривожланиш истикболлари” биринчи конференцияси тўплами. (4-қисм). - 27 апрель 2020 йил. – 602 б. - Б.581-585.
44. Хамроева Ш. Матнга ишлов бериш ва таҳлил қилиш дастурлари умумий шарҳи // “Ўзбекистонда илмий-амалий тадқиқотлар” мавзусидаги руспублика 19-қўп тармоқли илмий масофавий онлайн конференция материаллари тўплами. 31 август 2020 йил. – Тошкент: Tadqiqot, 2020. – 289 б. – Б. 82-83.
45. Хамроева Ш. Матнни автоматик таҳлил қилишнинг статистик воситалари ҳақида // “Ўзбекистонда илмий-амалий тадқиқотлар” мавзусидаги руспублика 19-қўп тармоқли илмий масофавий онлайн конференция материаллари тўплами. 31 август 2020 йил. – Тошкент: Tadqiqot, 2020. – 289 б. – Б. 84-85.
46. Хамроева Ш., Гулямова Ш. Роль электронных словарей в развитии компьютерной лингвистики // Инновационное развитие науки и образования. Международная научно-практическая конференция. – Казахстан. Павлодар. 2020 г, август. – Б. 88-93.
47. Xamroyeva Sh., Kuramshin R.M. O‘zbek tili morfologik analizatorini yaratishda “STARLING” morfologik analizatori tajribasidan foydalanish xususida // “Davlat tili – ijtimoiy taraqqiyot va milliy yuksalish mezon” mavzusida Respublika masofaviy ilmiy-amaliy anjumani ilmiy maqola va tezislari to‘plami. (2020-yil,16-oktabr). – Buxoro: 2020-y. 706 bet. – Б. 405-409.
48. Xamroyeva Sh., Gulyamova Sh. Elektron lug‘atlar – kompyuter leksikografiyasi mahsuli // “Davlat tili – ijtimoiy taraqqiyot va milliy yuksalish mezon” mavzusida Respublika masofaviy ilmiy-amaliy anjumani ilmiy maqola va tezislari to‘plami. (2020-yil,16-oktabr). – Buxoro, 2020. 706 bet. – Б. 459-465.
49. Xamroyeva Sh., Mengliyev B. O‘zbek tilshunosligida axborot-qidiruv tizimlari lingvistik ta’minoti borasidagi ishlar haqida // “Ona tili – millat ko‘zgusi” mavzusidagi respublika ilmiy-amaliy konferensiyasi materiallari. 19-oktabr 2020-yil. – Toshkent: O‘zbekiston xalqaro islom akademiyasi, 2020. – Б. 131-134.
50. Хамроева Ш., Менглиев Б. Туркий тиллардаги морфологик анализаторлар хусусида // Буюк ипак йўлида умуминсоний ва миллий қадриятлар: тил, таълим ва маданият”. Халқаро илмий-амалий конференция материаллари. – Самарқанд: СамДЧТИ, 2020. – 542 б. – Б. 35-39.
51. Xamroyeva Sh. Tilshunoslikdagi dolzarb yo‘nalishlar // “O‘zbek tilshunosligining dolzarb masalalari” mavzusidagi onlayn ilmiy anjuman materiallari / Mas’ul muharrirlar: Sh.Maxmaraimova, A.Eshmo‘minov. – Termiz, 2020. – 190 b. – Б. 17-18.
52. Mengliyev B., Khamroeva Sh. Linguistic support of semantic tags of noun units in uzbek language corpus // Диалектология. Этнолингвистика. Мифология. Ономастика. Этимология: материалы Международной научной конференции (Уфа, 1-4 ноября 2020). II том. – Уфа: Гилем, 2020. – 378 с. – С. 17-18.
53. Khamroeva Sh., Nabieva D. The importance of the corpus of special texts in the learning of written heritage // Диалектология. Этнолингвистика. Мифология. Ономастика. Этимология: материалы Международной научной конференции (Уфа, 1-4 ноября 2020). II том. – Уфа: Гилем, 2020. – 378 с. – С. 18-23.