

**ФАРҒОНА ДАВЛАТ УНИВЕРСИТЕТИ ҲУЗУРИДАГИ
ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.27.06.2017.Fil.05.02 РАҚАМЛИ ИЛМИЙ КЕНГАШ**

**ТОШКЕНТ ДАВЛАТ ЎЗБЕК ТИЛИ ВА АДАБИЁТИ
УНИВЕРСИТЕТИ**

АБЖАЛОВА МАНЗУРА АБДУРАШЕТОВНА

**ЎЗБЕК ТИЛИДАГИ МАТНЛАРНИ ТАҲРИР ВА ТАҲЛИЛ
ҚИЛУВЧИ ДАСТУРНИНГ ЛИНГВИСТИК МОДУЛЛАРИ
(РАСМИЙ ВА ИЛМИЙ УСЛУБДАГИ МАТНЛАР ТАҲРИРИ
ДАСТУРИ УЧУН)**

10.00.11 – Тил назарияси. Амалий ва компьютер лингвистикаси

**ФИЛОЛОГИЯ ФАНЛАРИ БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ**

Фарғона – 2019

**Филология фанлари бўйича фалсафа доктори (PhD) диссертацияс:
автореферати мундарижаси**

**Оглавление автореферата диссертации доктора философии (PhD) по
филологическим наукам**

**Contents of dissertation abstract of Doctor of Philosophy (PhD) on filological
sciences**

Абжалова Манзура Абдурашатовна

Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик
модуллари (расмий ва илмий услубдаги матнлар таҳрири дастури учун).....3

Абжалова Манзура Абдурашатовна

Лингвистические модули для программы редактирования и анализа текстов
на узбекском языке (для программы анализа официальных и научных
текстов).....23

Abjalova Manzura Abdurashetovna

Linguistic modules for the program of editing and analyzing texts in the Uzbek
language (for the program of analyzing official and scientific texts).....43

Эълон қилинган ишлар рўйхати

Список опубликованных работ

List of publications.....46

**ФАРҒОНА ДАВЛАТ УНИВЕРСИТЕТИ ҲУЗУРИДАГИ
ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.27.06.2017.Fil.05.02 РАҚАМЛИ ИЛМИЙ КЕНГАШ**

**ТОШКЕНТ ДАВЛАТ ЎЗБЕК ТИЛИ ВА АДАБИЁТИ
УНИВЕРСИТЕТИ**

АБЖАЛОВА МАНЗУРА АБДУРАШЕТОВНА

**ЎЗБЕК ТИЛИДАГИ МАТНЛАРНИ ТАҲРИР ВА ТАҲЛИЛ
ҚИЛУВЧИ ДАСТУРНИНГ ЛИНГВИСТИК МОДУЛЛАРИ
(РАСМИЙ ВА ИЛМИЙ УСЛУБДАГИ МАТНЛАР ТАҲРИРИ
ДАСТУРИ УЧУН)**

10.00.11 – Тил назарияси. Амалий ва компьютер лингвистикаси

**ФИЛОЛОГИЯ ФАНЛАРИ БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ**

Фарғона – 2019

Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида В2019.2.PhD/Fil.320 рақам билан рўйхатга олинган.

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

Диссертация автореферати уч тилда (ўзбек, рус, инглиз (резюме)) Фарғона давлат университети веб-сайти (www.fdu.uz) ҳамда «ZiyoNet» Ахборот-таълим порталида (www.ziyo.net) жойлаштирилган.

Илмий раҳбар:

Дадабоев Ҳамидулла Арипович
филология фанлари доктори, профессор

Расмий оппонентлар:

Улуқов Носиржон Муҳаммадалиевич
филология фанлари доктори, профессор

Пардаев Азамат Бахронович
филология фанлари доктори

Етакчи ташкилот:

Бухоро давлат университети

Диссертация ҳимояси Фарғона давлат университети ҳузуридаги филология фанлари бўйича илмий даражалар берувчи DSc.27.06.2017.Fil.05.02 рақамли Илмий кенгашнинг 2019 йил «___» _____ соат ___ даги мажлисида бўлиб ўтади (Манзил: 100151, Фарғона шаҳри, Мураббийлар кўчаси, 19-уй. Тел.: (99873) 244-66-02; факс: (99873) 244-44-01; e-mail: info@fdu.uz).

Диссертация билан Фарғона давлат университетининг Ахборот-ресурс марказида танишиш мумкин (___ рақами билан рўйхатга олинган). Манзил: 100151, Фарғона шаҳри, Мураббийлар кўчаси, 19-уй. Тел.: (99873) 244-71-28

Диссертация автореферати 2019 йил «___» _____ куни тарқатилди.
(2019 йил «___» _____ даги ___ рақамли реестр баённомаси)

М.Х.Ҳакимов

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

М.Т.Зокиров

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

С. Мўминов

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

КИРИШ (фалсафа доктори (PhD) диссертацияси аннотацияси)

Тадқиқот мавзусининг долзарблиги ва зарурати. Жаҳон тилшунослигида XX асрнинг 50-йилларидан эътиборан матнни электрон ҳисоблаш машиналари воситасида таҳрир қилиш муаммоси ўрганила бошлади, натижада замонавий ахборот технологиялари дастурлари ҳамда табиий тилнинг грамматик қонуниятларига асосланган матн таҳрир ва таҳлилларининг лингвистик таъминотини яратишга қаратилган салмоқли тадқиқотлар юзага келди. Бу асосда юз берган технологик тараққиёт дунёдаги етакчи адабий тилларда матнни квантитатив таҳрир ва таҳлил қилиш тизимларини яратишга олиб келгани маълум.

Дунё тилшунослигида автоматик таҳрирлаш йўналишида автоматик таҳрир қилувчи тизимларнинг лингвистик таъминотлари ишлаб чиқилган. Улар тил материални таҳрир ва таҳлил қилувчи тезкор ва иқтисодий тежамкор тизимларнинг яратилишига асос бўлиши баробарида машина таржимаси сифатини оширишда ҳам муҳим омил саналади. Зеро, глобаллашув ва жадал ривожланаётган фан-техника даврида катта ҳажмдаги ахборот устида тезлик билан ишлаш ва самарали натижа олиш жуда зарур. Шу маънода, дунё тилшунослиги фани олдида тилни лингвистик моделлаштириш, сўзларни леммалаш, таҳрир ва таҳлил алгоритмининг тузиш ҳамда лингвоанализ дастурининг лингвистик модулларини яратиш вазифаси қўйилмоқда.

Ўзбек тилшунослигида матнларни ахборот технологиялари дастурлари орқали танлаш ва уни қайта ишлаш бўйича бир қанча иш амалга оширилган бўлса-да, бу йўналишлардаги илғор жаҳон тажрибасини тадқиқотларда самарали қўллаш ҳамда матн таҳрир ва таҳлилларининг лингвистик таъминотини яратиш эҳтиёжи мавжуд. Зеро, "...эндиги вазифамиз – илмий тадқиқотларнинг натижаларини амалиётга татбиқ этишга қаратилган самарали механизмларни яратиш, илм-фанни, илмий-инновацион фаолиятни барқарор ривожлантириш" дир¹. Ҳозирги даврда дунё миқёсида матн билан боғлиқ технологияларнинг ривожланганлиги ва лингвистик дастурларнинг юзага келганлиги ўзбек тили учун яратиладиган автоматик таҳрир ва таҳлил (АТТ)нинг тараққиёти ва бугунги ҳолатини ўрганиш, матн таҳлили дастурининг лингвистик таъминотини тузишнинг умумий тамойилларини аниқлаш, бу хусусда янгича ёндашувлар асосида тадқиқотлар олиб бориш муҳим аҳамият касб этади.

Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги ПФ-4947-сон "Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида"ги, 2017 йил 17 февралдаги ПҚ-2789-сон "Фанлар академияси фаолияти, илмий тадқиқот ишларини ташкил этиш, бошқариш ва молиялаштиришни янада такомиллаштириш чора-тадбирлари тўғрисида"ги, 2016 йил 13 майдаги ПФ-4797-сон "Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетини ташкил этиш

¹ Мирзиёев Ш. Миллий тараққиёт йўлимизни қатъият билан давом эттириб, янги босқичга кўтарамиз. I жилд. – Тошкент: Ўзбекистон, 2017. – Б. 168-174.

тўғрисида”ги, 2018 йил 5 июндаги ПҚ-3775-сон “Олий таълим муассасаларида таълим сифатини ошириш ва уларнинг мамлакатда амалга оширилаётган кенг қамровли ислохотларда фаол иштирокини таъминлаш бўйича кўшимча чора-тадбирлар тўғрисида”ги Фармонлари ҳамда бошқа меъёрий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишга ушбу диссертация иши муайян даражада хизмат қилади.

Тадқиқотнинг республика фан ва технологиялари ривожланишининг устувор йўналишларига мослиги. Тадқиқот иши республика фан ва технологиялари ривожланишининг I. “Ахборотлашган жамият ва демократик давлатни ижтимоий, ҳуқуқий, иқтисодий, маданий, маънавий-маърифий ривожлантириш, инновацион иқтисодиётни шакллантириш” устувор йўналишига мувофиқ бажарилган.

Муаммонинг ўрганилганлик даражаси. Хорижда компьютер лингвистикаси соҳасида матнларни АТТ қилиш масаласи кўплаб олимлар томонидан ўрганилган². Ж.Петерсон хатоларни текширувчи ва уларни тузатувчи ДТлар борасида сўз юритган. Матннинг автоматик таҳлил жараёнини лингвистик аспектда ўрганиш Д.Юрафский номи билан боғлиқ. Олимнинг матнларни автоматик қайта ишлаш, автоматик таҳлил борасидаги тадқиқотлари маҳсули³ кейинги изланишларга тамал тошини қўйган.

Автоматик таҳрирлаш йўналишида автоматик таҳрир қилувчи дастурлар ишлаб чиқилган, улар русча, инглизча, немисча, французча, италянча матнларни таҳрирлашга йўналтирилган. Ҳатто татар, бошқирд, қозок, қирғиз, озарбайжон каби туркий тилларга мансуб матнлар учун морфологик ва синтактик анализаторлар яратилган, MS WORD системасига авто-таҳрир пакетлари жойлаштирилган⁴. Бундай таҳрирлаш дастурлари, асосан, имлони текширишга мўлжалланган. Рус тилини автоматик таҳлил қилиш йўналишида Р.Г.Котов, В.Е.Берзон, В.Г.Бритвин, И.А.Мельчук, А.Б.Кузнецов, Л.И.Беляева, В.А.Чижаковский, Г.Г.Белоногов, И.С.Дуганова, Л.Л.Иомдин, Э.К.Лавошникова, А.В.Сокирко, И.М.Ножов, Е.И.Большакова,

²Peterson, James. Computer Programs for Detecting and Correcting Spelling Errors. 1980.; Do Prado H.A. Emerging Technologies of Text Mining: Techniques and Applications / Ed. by H. A. Do Prado, E. Ferneda. Idea Group Reference, 2007. – 358 p.; Earnest, Les. The First Three Spelling Checkers. Stanford University. Archived from the original (PDF) on 22 October 2012.; Earnest, Les. Visible Legacies for Y3K (PDF). Archived from the original.; de Amorim, R.C.; Zampieri, M. Effective Spell Checking Methods Using Clustering Algorithms. Proceedings of Recent Advances in Natural Language Processing. Hissar, Bulgaria. 2013. – P. 172-178.; Zampieri, M.; de Amorim, R.C. Between Sound and Spelling: Combining Phonetics and Clustering Algorithms to Improve Target Word Recovery. Proceedings of the 9th International Conference on Natural Language Processing (PolTAL). Lecture Notes in Computer Science (LNCS). Springer. 2014. – P. 438-449.

³Jurafsky D., Martin J. Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. – New Jersey: PrenticeHall, 2000. – 927 б.

⁴Орехов Б.В., Слободян Е.А. Проблемы автоматической морфологии агглютинативных языков и парсер башкирского языка [Текст] // Информационные технологии и письменное наследие: материалы международной научной конференции (Уфа, 28-31 октября 2010 г.) / отв. ред. В. А. Баранов. – Уфа. Ижевск: Вагант, 2010. – С. 167-171.; <http://lcp.bashedu.ru/index.php?go>. Орехов Б., Галлямов А. Башкирский морфологический анализатор (Электронный ресурс); <http://my.ict.nsc.ru/~salerat/kaz/>. Программа генерации словоформ казахского языка; <http://tamgasoft.kg/morfo/ru/>. Морфология кыргызского языка; Сулейманов, Д. Ш. Синтаксический анализатор предложений татарского языка / Д. Ш. Сулейманов, А. Р. Гатиатуллин // Труды Математического центра имени Н. И. Лобачевского. – Казань, 1999. – Т. 4. – С. 111-126.

Д.В.Клышинский, Н.Н.Леонтьева⁵ ва бошқалар самарали тадқиқот олиб боришган.

Туркий тилларга мансуб матнларни статистик методлар асосида лингвистик таҳлил қилиш соҳасида Р.Г.Пиотровский, М.А.Махмудов, А.А.Пиотровская, Т.Садиқов ишлари диққатга сазовор⁶. Ўзбек тили материаллари бўйича КЛга оид дастлабки тадқиқотларни олиб борган олимлар Ҳ.Арзиқулов, С.Ризаев ва С.Мухамедовлардир⁷. Улар, асосан, КЛнинг статистик таҳлил йўналишлари бўйича назарий ва амалий ишларни амалга оширган.

XXI аср бошига келиб ўзбек ва инглиз тилларида бир қатор изланишлар қилган, ўзбек КЛ ривожига салмоқли улуш қўшган А.Пўлатов бошчилигида ўзбек тилшунослигига оид масалаларни компьютер воситасида ҳал этишга бағишланган битирув малакавий ишлари, магистрлик диссертациялари ҳимоя қилинди, ўнлаб илмий мақолалар чоп этилди, ўқув қўлланмалар яратилди⁸ ва китоблар нашрдан чиқарилди⁹. Айниқса, феъл туркумини формаллаштириш, уларнинг моделини ишлаб чиқиш ва феъллар асосида лингвистик таъмин яратишга бағишланган талай ишлар дунё юзини кўрди¹⁰. Сўнгги йилларда

⁵Котов Р.Г. Лингвистические аспекты автоматизированных систем управления. – Москва: Наука, 1977. Он же. Прикладная лингвистика и информационная технология. – Москва: Наука, 1987.; Котов Р.Г., Якунин Б.В. Язык информационных систем. – Москва: 1989.; Он же. Оптимизация речевого воздействия. – Москва: Наука, 1990.; Берзон В.Е. и др. О разработки подсистемы редактирования выходного текста в рамках системы МП // Проблемы внутренней динамики речевых норм. – Минск, 1992, – С. 169-174; Бритвин В.Г. Прикладное моделирование синтагматической семантики научно-технического текста (на примере автоматического индексирования): Дисс. канд. филол.наук. – Москва: МГУ, 1983.; Мельчук И.А. Порядок слов при автоматическом синтезе русского слова (предварительные сообщения)/ Научно-техническая информация. 1985, №12. – С. 12-36.; Беляева Л.И., Чижаковский В.А. Тезаурус в системах автоматической переработки текста. – Кишинев, 1983.; Белоногов Г.Г., Котов Р.Г. Автоматизированные информационно-поисковые системы. – Москва, 1968; Белоногов Г.Г., Дуганова И.С., Кузнецов А.Б. Экспериментальная система автоматизированного обнаружения и исправления ошибок в тексте / НТИ. Серия 2, 1984, – №3. – С. 20-25.; Иомдин Л.Л. Автоматическая обработка текста на естественном языке: модель согласования. – Москва: Наука, 1990. – 180 с.; Лавошникова Э.К. О «подводных камнях» в компьютерных системах проверки правописания // Вестник Московского университета. Серия 9. Филология. 2002, – №6. – С.151-163.; Сокирко А.В. Семантические словари в автоматической обработке текста: По материалам системы ДИАЛИНГ: Дисс. ... канд.тех.наук. – Москва, 2001. – 120 с.; Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы) сегментации русского предложения: Автореф. дисс. ...канд. филол.наук. – Москва, 2003.; Большакова Е.И., Клышинский Э.С. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб.пособие. – Москва: МИЭМ. 2011. – 272 с.; Леонтьева Н.Н. Автоматическое понимание текста: системы, модели, ресурсы: учеб.пособие. – Москва: Академия. 2006. – 304 с.

⁶Махмудов М.А., Пиотровская А.А., Садыков Т. Система системного анализа и синтеза тюркской словоформы / Переработка текста методами инженерной лингвистики. – Минск, 1982.; Пиотровский Р.Г. и др. Методы автоматического анализа и синтеза текста. – Минск: Высшая школа, 1985.

⁷Арзиқулов Х.А., Пиотровская К.Р. Информатика и переработка текста средствами вычислительной техники (учебное пособие). – Самарканд, 1986.; Ризаев С. Ўзбек тилининг лингвостатистик тадқиқи: Филол. фан. док. ...дисс. автореф. – Тошкент, 2008. – 50 б.; Шу муалл. Ўзбек тилшунослигида лингвостатистика муаммолари (монография). – Тошкент: Фан. 2005. – 295 б.; Мухамедов С.А. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дис. ...канд.филол.наук. – Тошкент, 1980. – 25 с.

⁸Мухамедова С. Компьютер лингвистикаси (методик қўлланма). – Тошкент, 2007.; Пўлатов А., Мухамедова С. Компьютер лингвистикаси (ўқув қўлланма). – Тошкент, 2008. – 98 б.

⁹Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011. – 520 б.; Rahimov A. Kompyuter lingvistikasi asoslari. – Toshkent: Akademnashr. 2011. – 160 b.; Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.

¹⁰Пўлатов А.Қ., Алиходжаев Б., Джураева Н. Разработка программы компьютерного анализа и синтеза глаголов узбекского языка // ЎзМУ хабарлари. – Тошкент, 2002. – №2. – С. 17-19.; Po'latov A.Q., Mo'minova

М.Х.Хакимов раҳбарлигида “Таржимон Л-МХ” системаси учун бажарилган тадқиқот ишларида ўзбек ва рус тилларининг лексик-синтактик таҳлили берилган¹¹. Н.Абдурахмонова, Ш.Ҳамроева ва Д.Ўринбоева¹²лар изланишларида матнларни таҳлил қилиш бўйича муайян фикрлар берилган, бу жиҳат лингвистик дастурларни яратишда лингвистик модулларни ишлаб чиқишнинг муҳим аҳамиятга эгалигини белгилаб берган.

Юқорида тилга олинган тадқиқотчиларнинг илмий изланишлари натижаси ўлароқ матнларни АТТ қилишга доир етарлича назарий билим ҳамда муайян тажрибалар тўпланди. Уларга асосланган ҳолда мазкур тадқиқотда дастлабки ташаббус – ўзбек компьютер лингвистикаси (КЛ)да автоматик таҳрирловчи дастурни яратиш учун ҳаракат қилдик. Матнларни автоматик таҳрир қилувчи дастурлар лингвистик модулларининг яратилиши мукамал дастурлар ишлаб чиқилишига замин яратади, бу эса ўзбек адабий тилидаги матнларнинг саводли тайёрланишига хизмат қилади.

Тадқиқотнинг диссертация бажарилган олий таълим муассасасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги. Диссертация тадқиқоти Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университети илмий-тадқиқот ишлари режасининг “Тилнинг ижтимоий, тарихий ва замонавий тараққиётини ўрганиш” йўналиши доирасида бажарилган.

Тадқиқотнинг мақсади ўзбек адабий тилининг расмий ва илмий услубларидаги матнларни графематик таҳрир ва морфологик, синтактик таҳлил қилувчи дастурнинг лингвистик модулларини яратиш ҳамда ўзбек тилининг луғавий-мажмуий бойлигини ўзида мужассам қилувчи лингвистик таъминотни шакллантиришдан иборат.

Тадқиқот вазифалари:

матнларни компьютер ёрдамида таҳрир ва таҳлил қилиш жараёнини ва машина таҳлилини амалга оширувчи илғор технологияларнинг мазмун ва моҳиятини аниқлаш ва таснифлаш;

ўзбек тили сўз туркумлари модулида лексик-грамматик кодни шакллантириш мақсадида фонологик, морфонологик ва орфографик қоидалар алгоритмини тузиш; синтактик ва семантик кодни шакллантирувчи сўз бирикмаларининг лингвистик моделини яратиш;

автоматик лингвистик таҳлил босқичларини ўрганиш ва шу асосда таҳрир қилувчи дастур босқичлари модулларини ишлаб чиқиш, лингвистик

Т., Po'latova I.O. Dunyoviy o'zbek tili (O'zbek tilida fe'lining shakllari va ularning rus, ingliz tillarida berilishi). – Toshkent: Universitet, 2003. – 404 b.; Мухамедова С. Ўзбек тилидаги ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. – Тошкент, 2006. – 80 б.

¹¹Хакимов М.Х. Расширяемый входной язык математического моделирования естественного языка для много-язычной ситуации машинного перевода // ЎЗМУ хабарлари, 2009. – № 1. – С.75-80.

¹²Абдурахмонова Н. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти: Филол.фан.бўйича фалсафа доктори (PhD)...дис. афтореф. – Тошкент, 2018. – 47 б.; Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол.фан.бўйича фалсафа доктори (PhD)...дис. афтореф. – Қарши, 2018. – 53 б.; Ўринбоева Д. Халқ оғзаки ижоди: жанрий-лисоний ва лингвостатистик тадқиқ муаммолари: Филол. фан. бўйича докт. (DSc) ... дисс. автореф. – Самарқанд, 2019. – 74 б.

тахрирлаш ва таҳлил қилиш дастурини яратишдаги муаммоларни аниқлаш, уларнинг ечимини бериш;

дастурнинг лингвистик таъминотини шакллантириш йўлларининг оптимал вариантларини яратиш ҳамда модуллар ёрдамида дастурнинг лингвистик алгоритминини тузиш.

Тадқиқотнинг объекти учун ўзбек адабий тилининг расмий-идоравий ва илмий услубларида ёзилган матнлар танланди.

Тадқиқотнинг предмети ўзбекча матнларнинг автоматик таҳрир ва таҳлил дастури учун лингвистик таъминот асосларини ишлаб чиқиш.

Тадқиқотнинг усуллари. Мавзуни ёритишда тавсифий, компонент таҳлил, статистик, чоғиштирма-типологик, моделлаштириш, шажаравий синтактик таҳлил, аналогия методларидан фойдаланилди.

Тадқиқотнинг илмий янгилигини қуйидагилар ташкил этади:

ўзбек компьютер лингвистикасида автоматик таҳрир ва таҳлил тизими лингвистик таъминотини шакллантиришнинг назарий асослари яратилган;

автоматик таҳрир ва таҳлил лингвистик таъминотини тузишда сўзларни морфем валентлик, туркумлик, синтагматик ва услубий белгиси бўйича ажратиш тамойиллари ишлаб чиқилиб, лингвистик модуль ва модуллашнинг аҳамияти асосланган;

расмий ва илмий матнлар бўйича ўзбек тилининг грамматик категориялари таъминоти, сўз туркумлари бўйича гуруҳланиши, жумла таркибида битишув, мослашув ва бошқарув йўли билан бирикиш моделлари шакллантирилган;

автоматик таҳрир/таҳлил қилиш технологияси ва дастури мукамаллигини таъминловчи графематик таҳрир, морфологик ва синтактик таҳлил модуллари ишлаб чиқилган.

Тадқиқотнинг амалий натижалари қуйидагилардан иборат:

ўзбек тилида мавжуд омоним сўз ва фразеологик бирликлар тўпланиб, базаси шакллантирилган, вариантдош сўз ва тасвирий ифодалар ҳамда уларни фарқлаш омиллари аниқланган, ўзбек тилидаги сўз ясовчи аффикслар статистикаси аниқланган;

ўзбекча матнлардаги омоним шаклларни таҳлил қилиш технологияси яратилган;

матнлардаги лексемаларни текшириш формуласи, омонимларни таҳлил қилиш технологияси яратилган;

АТТ лингвистик таъминотининг лексикографик манбалари, дастурий таъминот учун лингвистик база яратиш тамойиллари ва морфологик таҳлил алгоритми тузилган.

ўзбекча матнларни АТТ қилиш дастури ишлаб чиқилган.

Тадқиқот натижаларининг ишончлилиги тилшунослик ва КЛда назарий асосланган илмий хулосалар, бошқа тиллардаги мавзуй нашрлар, ишнинг амалий аҳамиятининг долзарблиги, шунингдек, “Ўзбек тилининг изоҳли луғати”, “Ўзбек тили омонимларининг изоҳли луғати”, “Ўзбек тилининг асосий имло қоидалари”, “Ҳозирги ўзбек адабий тили”

китобларидаги қарашларда ўз ифодасини топган хулосаларга мос келиши билан белгиланади.

Тадқиқотнинг илмий ва амалий аҳамияти. Матнларни АТТ қилиш борасидаги лингвистик масалаларнинг тадқиқ этилганлиги КЛнинг муайян йўналишларига доир методологик назарияларни таҳлил қилиш, изланишлар доирасини янада кенгайтириш, лингвистик таъминот (ЛТ) модулларини яратиш, таълим муассасаларида ўзбек тилига ўқитиш жараёнини олиб бориш учун хизмат қилади. Тадқиқот жараёнидаги кузатиш ва илмий-назарий хулосалардан олий ўқув юртлари, ўрта махсус таълим тизимлари учун курс ва семинарларда фойдаланиш мумкин. Тадқиқот ўзбек тилидаги матнларни АТТ қилиш билан боғлиқ монография, дарслик ва қўлланмалар яратишда илмий аҳамиятга эга.

Мазкур тадқиқотда ишлаб чиқилган лингвистик модул (ЛМ)лар ва уларнинг алгоритмлари натижасида ўзбек тилининг расмий ва илмий услубларидаги матнларни таҳрир қилувчи дастурнинг яратилишига эришилади. Шунингдек, матнларни қайта ишлаш дастурий таъминотларининг лингвистик базасини яратишда фундаментал манба вазифасини бажаради. Компьютер лингвистикаси, прагмалингвистика, когнитив тилшуносликнинг муайян масалаларига оид курслар, семинарлар ташкил қилиш ва амалий машғулотлар учун ўқув қўлланмалари яратишга хизмат қилади. Бу жиҳат тадқиқот ишининг амалий аҳамиятини белгилаб беради.

Тадқиқот натижаларининг жорий қилиниши. Диссертацияда расмий ва илмий услублардаги ўзбекча матнларни таҳрир қилувчи дастурни яратиш бўйича олинган натижалар қуйидаги йўналишларда жорий қилинган:

сўз ясалишининг аффиксация усули лексика ва статистикаси ҳамда ўзбек тилига формал ёндашиш натижаларидан “Ўзбек тили ва унинг таржимаси билан боғлиқ электрон ва банк-молия терминларининг кўп тилли луғатларини яратиш” инновацион тадқиқот лойиҳасида фойдаланилган (Олий ва ўрта махсус таълим вазирлигининг 2019 йил 5 мартдаги 89-03-915-сон маълумотномаси). Илмий натижаларни жорий қилиш асосида кўп тилли луғатлар мазмуни мукаммаллашишига эришилган;

АТТ дастурлари ёрдамида синтактик таҳлил қилиш бўйича тадқиқ хулосалари ҳамда сўз бирикмалари моделларидан “Кўриш имконияти чекланган талабалар учун тилшунослик фанлари бўйича аудиокитоб таъминотини яратиш” мавзусидаги давлат амалий гранти лойиҳасида фойдаланилган (Олий ва ўрта махсус таълим вазирлигининг 2019 йил 5 мартдаги 89-03-915-сон маълумотномаси). Натижада “Ҳозирги ўзбек адабий тили” фанининг синтаксис бўлимига оид лаборатория топшириқлари тизими ишлаб чиқилган;

тадқиқот натижаларидан Ўзбекистон Миллий телерадиокомпанияси тизимидаги “Navoiy” телеканалининг “Дахлдорлик туйғуси” кўрсатуви ва “O‘zbekiston” радиоканалининг “Тил – миллат кўзгуси”, “Таълим ва ислоҳот” радиоэшиттиришлари сценарийларини тайёрлашда фойдаланилган (Навоий вилояти телерадиокомпаниясининг 2019 йил 11 мартдаги 01-02/57-сонли

маълумотномаси). Натижада кўрсатув ва радиоэшиттиришларнинг илмий-оммабоплиги таъминланган.

Тадқиқот натижаларининг апробацияси. Тадқиқот натижалари 6 та халқаро, 7 та республика ва халқаро миқёсдаги илмий-амалий анжуманда муҳокамадан ўтказилган.

Натижаларнинг эълон қилинганлиги. Диссертация мавзуси бўйича жами 31 та илмий иш чоп этилган, жумладан, 1 та илмий гувоҳнома, Ўзбекистон Республикаси Олий аттестация комиссиясининг докторлик диссертациялари асосий илмий натижаларини чоп этиш тавсия этилган илмий нашрларда 5 та мақола, жумладан, 3 таси республика ҳамда 2 таси хорижий журналда эълон қилинган.

Диссертациянинг тузилиши ва ҳажми. Диссертация кириш, уч боб, хулоса, фойдаланилган адабиётлар рўйхатидан иборат, ҳажми – 164 саҳифа.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

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

Диссертациянинг **“Когнитив тадқиқотлар парадигмасида матн таҳрири ва таҳлили”** деб номланган биринчи бобида ушбу йўналишда ва когнитив билимлар соҳасида табиий тилни тушуниш учун компьютер моделларини яратиш мақсадида олиб борилган кўплаб тадқиқотлар таҳлил қилинган.

Матн таҳлилининг дастурий таъминотини яратиш бўйича рус компьютер лингвистикасида самарали изланишлар олиб борилган. Хусусан, КЛда автоматик таҳлилнинг умумий масалалари, моделлаштириш принциплари¹³, морфологик таҳлил¹⁴, синтактик¹⁵ ва семантик таҳлил¹⁶,

¹³Ягунова Е.В. Основы теоретической, вычислительной и экспериментальной лингвистики / Е.В.Ягунова // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие / Е. И. Большакова [и др.]. – Москва: МИЭМ. 2011. – 272 с. – С. 7-89.

¹⁴Клышинский, Э. С. Начальные этапы анализа текста / Э. С. Клышинский // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие / Е. И. Большакова [и др.]. – Москва: МИЭМ. 2011. – 272 с. – С. 106-140.; Марчук, Ю. С. Компьютерная лингвистика / Ю. С. Марчук. – Москва, 2007. – С. 60-70.

¹⁵Компьютерный синтаксический анализ: описание моделей и направлений разработок / Г. Д. Карпова [и др.] // Итоги науки и техники. Серия «Вычислительные науки». Т. 6. – Москва, 1991. – 243 с.; Иорданская, Л. Н. Автоматический синтаксический анализ. Т. 2. Межсегментный синтаксический анализ / Л. Н. Иорданская. – Новосибирск, 1967. – 231с.; Мельчук, И. А. Автоматический синтаксический анализ. Том 1. Общие принципы. Внутрисегментный синтаксический анализ. / И. А. Мельчук. – Новосибирск, 1964. – 360 с.

¹⁶Апресян, Ю.Д. Исследования по семантике и лексикографии. Т. I: Парадигматика. – Москва: Языки славянских культур, 2009.; Сокирко А.В. Семантические словари в автоматической обработке текста: По материалам системы ДИАЛИНГ: Дисс. ... канд.тех. наук. – Москва, 2001. – 120 с.

таҳлил бирликлари, табиий тилдаги матнларни автоматик тушуниш назарияси¹⁷, матнни таҳлил қилиш босқичлари, морфологик анализ ва синтез, табиий тилдаги матнни қайта ишлашга бағишланган бир қатор асарлар майдонга келди. Бироқ автоматик таҳрир ва таҳлил жараёни ўзбек КЛда ўрганилмаган.

Бугунги кунда кенг тарқалган инглиз тилидаги матнларни таҳрир ва таҳлил қилиш дастурларининг яратилиш тарихини америкалик компьютер технологиялари олими Лестер Дональд Эрнест бошлаб берган. 1971 йил февраль ойида Л.Эрнест раҳбарлигида Стенфорд университетининг сунъий интеллект лабораториясида аспирант Ральф Горин (DEC PDP-10 учун SPELL (имло текшириш)) инглиз тилидаги матнлар учун илк имло текширувчи ишланма эмас, балки амалий дастурни яратади¹⁸. Горин SPELLнинг тезроқ ишлаши учун дастурни ассемблер дастурлаш тилида ёзган. Шунингдек, диссертацияда рус тилидаги матнларнинг автоматик грамматик таҳлилини амалга оширувчи ОРФО системаси яратувчиси “Информатик” компаниясининг лингвистик модуллари кўрсатилди.

Ўзбек тилидаги матнларни (биринчи қадамда) Microsoft Office Word дастурига киритиш жараёнида уларни автоматик таҳрир ва таҳлил қилишга эришиш, бунинг учун дастурнинг лингвистик таъминини яратиш мақсадида бугунги кунда кенг қўлланишда бўлган дастурлар чуқур ўрганилди. Уларнинг амалий аҳамияти қисқа шарҳ тарзида берилди. Шунингдек, матнларни автоматик таҳрир ва таҳлил қилиш назарий жиҳатдан ўзбек тилшунослигида илк бор тадқиқ қилинаётгани боис жаҳон КЛдаги мазкур иш билан боғлиқ терминларга изоҳлар қайд этилди.

Хориж компьютер лингвистикасида фаол истеъмолда бўлган *модуль* термининг моҳияти, бошқа соҳаларда қўлланиш кўлами ёритилди. Лингвистик модуль лингвистик дастурларнинг *мустақил таркибий қисмлари*¹⁹, яъни дастурий таъминотнинг муайян лингвистик жараёнини қамраган қисми ҳисобланади. Лингвистик таъминот, бошқача айтганда, лингвистик процессор ушбу қисмлар мажмуини ташкил этади. Лексик модулда луғат қатлами қамраб олинса, графематик модулда рамзлар, тиниш белгилар, ҳарфий ва бошқа белгилар таҳрир қилинади, орфографик модулда имло қоидалари, морфологик модулда сўзшакллар анализи (сўзшаклдан лексемага қадар таҳлил) ва синтези (лексеманинг грамматик шаклланиши таҳлили жараёни), синтактик модулда суперсинтактик бирлик – гап ёки сўзларнинг ўзаро боғланиш ҳодисаси таҳлил қилинади.

Лотин ёзувига асосланган янги алифбони ислоҳ қилиш масаласи кун тартибида тургани боис мазкур тадқиқотда кирилл ёзувига асосланган имло қоидалари ва лексик манбаларига таянилди.

¹⁷Леонтьева Н.Н. К теории автоматического понимания естественных текстов. – Москва: МГУ, 2000. – 304.

¹⁸Peterson, James. Computer Programs for Detecting and Correcting Spelling Errors. 1980. https://en.wikipedia.org/wiki/Spell_checker

¹⁹ Демьянков В.З. Когнитивная лингвистика как разновидность интерпретирующего подхода // Вопросы языкознания. – Москва, 1994. – № 4. – С. 22.

Ўзбек тилидаги матнларни автоматик таҳрир ва таҳлил қилиш дастурининг лингвистик процессорини яратишда тилшуносга ўзбек адабий тилининг лингвистик луғатлари ва грамматик қоидалар жамланмаси зарур бўлади. Мазкур тадқиқот жараёнида лингвистик таъминотда грамматик шаклланган бирикмалар тарзида 105 перефрази (тасвирий ифода), 1393 фразема базаси яратилди. Шунингдек, матнда мавжуд омонимшакллар таҳлилни амалга оширишни таъминлаш мақсадида 1378 омонимшакл жамланди.

Е.И.Большакова ўз изланишларида лингвистик процессор лингвистик маълумотлар моделига асосланишини, яъни дастурий таъминот (ДТ)нинг лингвистик ресурсларини қуйидагилар ташкил этишини таъкидлайди: компьютер луғатлари, табиий тил грамматикаси, сўз бирикмалари базаси, тезаурус ва онтологиялар, матнлар мажмуаси ва корпус²⁰.

Лингвистик таъминоти таркибининг тўлдириб борилишига кўра АТТ дастурини *турғун ва динамик* (ўзгарувчан) турларга ажратиш мумкин. Турғун таҳрир дастури лингвистик таъминоти фақат мутахассис томонидан яратилади, унинг неологизмлар, янги лингвистик моделлар, тил меъёрлари билан боғлиқ ўзгаришлар дастурнинг кейинги версияларига киритиб борилади. Динамик муҳаррир дастурининг вазифаси доимий равишда лексикографик таъминотни янги сўзлар билан тўлдириб, бойитиб бориш ҳисобланади. Кўп ҳолларда ушбу вазифа фойдаланувчи томонидан амалга оширилади, яъни матнда қўлланилаётган муайян неологизм, янги термин махсус сўровларга биноан таъминотнинг зарур модулига киритилиши натижасида кейинги ўринларда унинг турли шаклларда қўлланилиши таҳрир ва таҳлил қилинишига эришилади. Бу матн муҳаррир дастури ривожини учун хизмат қилади, чунки тил доимий ҳаракатда, сўзлар янги-янги маънолар касб этиши, неологизм, архаизм, историзмларнинг пайдо бўлишини кузатиб бориш имконини беради. Е.И.Большакова бундай имкониятга морфологик луғатни бойитиш турлари сифатида қарайди²¹ ва уни икки турга ажратади: 1) луғатли бойитиш (турғун ДТ); 2) луғатсиз бойитиш (динамик ДТ) имконияти.

Матнларни АТТ қилиш дастурини инновацион таълимнинг бир кўриниши сифатида қайд этиб, унинг қуйидаги имкониятларга эга эканлигини кўрсатиш мумкин:

- 1) турли ҳажмдаги матнларни тезкор таҳрир қилади;
- 2) матн терувчининг ёзма нутқ малакасини такомиллаштиради;
- 3) фойдаланувчига матнларни мустақил равишда тўғри ва саводли ёзишида кўмак беради;
- 4) ўзбек тилининг лексик, грамматик меъёрларини ўрганишда лингвистик манба вазифасини ўтайди.

²⁰Большакова Е.И.и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие. – Москва: МИЭМ, 2011. – С. 97-99.

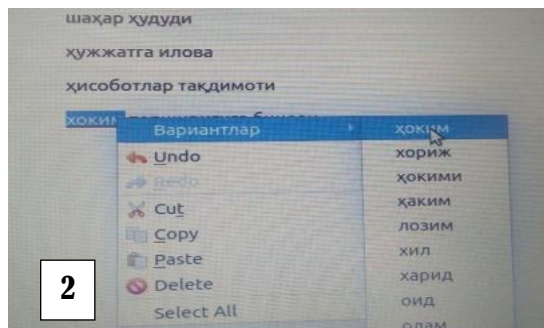
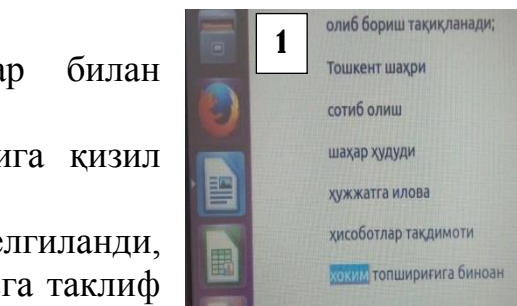
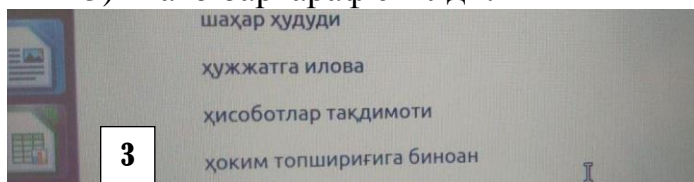
²¹Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие. – Москва: МИЭМ, 2011. – С. 112-122.

Фикримизга дастурдаги фрагментлар билан аниқлик киритамиз:

1) “*хоким*” тарзида ёзилган сўз тагига қизил тўлқинли чизик чизилди;

2) тагига чизилган сўз белгиланди, сичқончанинг ўнг томони босилиб хато сўзга таклиф қилинаётган сўзшакллардан керакли сўз танлаб олинди;

3) хато бартарарф этилди.



Эътиборли жиҳати шундаки, дастур томонидан фойдаланувчига таклиф қилинаётган сўзлар қаторида *хоким* лексемаси билан боғлиқ сўзшакллар учрамади. Бу жиҳат фойдаланувчига мазкур сўзнинг имло луғатида мавжуд эмаслиги ва кейинги ҳолатларда шу сўз билан боғлиқ хатога йўл қўймаслик кераклиги борасида хулосани беради. Бу дастурнинг бевосита таълимий аҳамиятини намоён этади.

Диссертациянинг “**Маттни автоматик таҳрир ва таҳлил қилиш системаси модуллари**” деб номланган иккинчи бобида матнларни таҳрир ва таҳлил қилувчи дастурнинг мукамал бўлиши учун қуйидаги лингвистик модулларни яратиш тамойиллари таҳлил қилинди: графематик таҳрир модули – матнда қўлланган турли рамзлар, белгилар ва алоҳида қўлланган ҳарфларни таҳрир қилиш босқичи²²; морфологик таҳлил модули – сўзшаклдан унинг лемма (лексеманинг луғатдаги шакли)га қадар таҳлил қилиниши; синтактик таҳлил – матндаги сўзларнинг грамматик боғланиши аниқланади.

Графематик таҳрирлаш (ГТ) босқичида матнда қўлланилган элементлар оқимини сўзшакл ва турли рамзларга ажратиш, маттни морфологик таҳлил жараёнига тайёрлаш, маттни корпусга киритиш мақсадида *токенизация* технологиясидан фойдаланиш мақбул ҳисобланади. Боиси токенизация электрон матндаги барча бирликларни ажратиш жараёнидир. Буни қуйидаги мисол билан кўрсатиш мумкин:

[‘Ўзбекистон’, ‘Республикаси’, ‘Президентининг’, ‘2005’, ‘йил’, ‘15’, ‘июндаги’, ‘«’, ‘Шарқ’, ‘тароналари’, ‘»’, ‘халқаро’, ‘мусиқа’, ‘фестивалига’, ‘тайёргарлик’, ‘кўриш’, ‘ва’, ‘уни’, ‘ўтказиш’, ‘тўғрисида’, ‘», ‘зи’, ‘ПҚ’, ‘-’, ‘103’, ‘-’, ‘сонли’, ‘қарори’, ‘ўз’, ‘кучини’, ‘йўқотган’, ‘деб’, ‘ҳисоблансин’, ‘.’]

Бу ерда квадрат кавс матннинг бошланиш ва яқунланиш жойини кўрсатади. Биртирноқ элементни ажратувчи белги, вергул ажратилган белгиларни бир-биридан алоҳидалашга хизмат қилади. Мазкур мисолда токенизация жараёнида ‘зи’ бўғини маънога эга бўлмагани боис кесиб

²² Леонтёва Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы: учебное пособие. – Москва, 2006. – С.50.

ташланади. Мазкур жараёндан сўнг элементларни теглаштириш, уларни ҳисоблаш, гуруҳлаштириш, ортиқча қўлланган белги ёки маънога эга бўлмаган сўзларни ўчириш мумкин.

Лингвистик таҳрир дастурининг ГТ босқичи модулларини яратишда матн элементлари типлари юзага чиқарилади²³. Бундай типлар қўйидаги бирликларни ўз таркибига олади: ЛекБ – лексик бирлик, бир ёзувга мансуб алифбодаги ҳарфлардан юзага келган лексема (*valida, buyruq* каби); ЧетЛ – чет тили лексемаси (*принтер*); РБ – рақамли бутунлик (1986, 18/04/2012, 5.05, 19,25); ХРБ – ҳарф-рақамли бутунлик (*Боинг-767, “Йил аёли – 2019”*); АББР – аббревиатура (*МДХ, ЎЗР, ДХХ, ЎЗМУ*); ҚБ – қисқартирилган бирикмалар (*ва ҳ.к., ва бошқ., м-н:*).

Морфологик таҳлил (МТ) тарихи ва унинг моҳияти ёритилди. КЛда ўтган асрнинг 60-70-йилларига келиб, машинали морфология қамровидаги барча тадқиқотлар машина луғатини яратиш билан бошланган²⁴.

МТда, асосан, имловий хатоларни текшириш кўзда тутилади ва асосий эътибор сўзшаклларга қаратилади. Ушбу мақсадга эришиш учун қўйидаги ишлар амалга оширилиши талаб қилинади: 1) тилнинг луғат таркибини компьютерга киритиш, яъни электрон луғатни тузиш; 2) луғатдаги адабий тилга мансуб сўзларни ажратиш олиш; 3) ажратиш олинган сўзларни услубий (илмий, бадиий, публицистик, расмий-идоравий ва сўзлашув) жиҳатдан гуруҳлаш; 4) адабий сўзларни туркумларга ажратиш; 5) сўзларни туркум доирасида гуруҳлаш (масалан, от туркумидаги сўзларни шахс отлари, ўсимлик ном (фитоним)лари, жой ном (топоним)лари каби гуруҳларга ажратиш); 6) гуруҳланган сўзларнинг аффикслар комбинациясини тузиш; бунда аффикслар комбинациясининг амалда адабий тил доирасида фойдаланиладиган ҳажми қамраб олинади. Комбинациялар қаторини тузишда аффикслар кетма-кетлигига эътибор қаратилади. М-н, от туркумида *кўплик+эғалик +келишик* категорияси ва ҳ.к. Грамматик воситанинг сўз таркибидаги тартиби маълум қонуниятга эга бўлиб, грамматик воситанинг жойлашишидаги тартиб ва изчиллик унинг маъно ва грамматик хусусияти билан боғлиқ: *янги лугавий маъно ҳосил қилувчи восита биринчи* (1), *лугавий маънога таъсир қилувчи восита иккинчи* (2) ва *лугавий маънога таъсир қилмайдиган, лекин сўзни боғлайдиган восита* (3) учинчи бўлиб қўшилади: S+1+2+3.

Синтактик шаклнинг тартиби исмларда: эғалик аффикси+келишик аффикси: *мартабасини, ўқиганимни*; феълларда: тасдиқ-инкор+замон/майл+шахс+сон: *ўтмаяпсан*. Ҳамма турдаги аффиксдан кейин аффикс-юклама қўшилади: *давлатимизними, лойиҳамгами, ўқимаяпсанми*.

7) Тузилган аффикслар комбинацияси сўзларга бириктирилади; 8) тилшунослар томонидан амалга оширилган юқоридаги ишлар дастурчи

²³ Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы: учебное пособие. – Москва: Академия, 2006. – С.58.

²⁴ Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы): Дис. канд. филол.наук. – Москва, 2003. – С. 54.

томонидан таъминотга киритилади.

Жаҳон КЛда МТ технологиялари яратилган бўлиб, бугунги кунда стемминг ва лемматизация самарадор технологияларга айланган. Агглютинатив табиатга эга ўзбек тилидаги матнларни АТ қилишда стемминг технологиясидан унумли фойдаланиш мақсадга мувофиқ.

Маълумки, эгалик ва келишик аффикслари билан ўзгариш хусусиятига эга сўзлар исмлар атамаси остида бирлаштирилади. Улар таркибига от, сифат, сон, олмош, тақлид сўз, феълнинг сифатдош ва ҳаракат номи киради. Шунинг назарда тутиб, тилшуносликдаги турланиш ва бошқа грамматик категорияларнинг сўзларга бирикиб келиш ҳодисаси исмларга мансуб туркумларда кузатилгани боис дастурнинг ЛТни яратишда морфологик синтез исмлар доирасида амалга оширилди. Бунинг учун луғавий шакл ясовчи ва синтактик шакл ҳосил қилувчи аффикслар базаси Accessда яратилди (gr_form). Кўплик шаклига эга ёхуд семантик кўплика эга сўзлар (асосан, саналмайдиган отлар, мавхум отлар) sol_CorrNounNumber гуруҳига бирлаштирилди.

Ўзбек тилидаги исмларнинг грамматик шаклланиши учун махсус дастурий қисм (sol_GenIsmForm) яратилди. Бунда 1) грамматик шакллар номи жадвалга махсус белги бериб киритилди (id_entry,iform); 2) id_dims майдонида эга аффикслар ID_Wordга боғланади, натижада List of derivate майдонида сўзшаклни юзага келтиради.

```
SELECT sg_form.name
FROM sg_entry, sg_form, coord_pairs
WHERE sg_entry.name='давлат'
AND sg_form.id_entry=sg_entry.id
AND coord_pairs.id=sg_form.id_dims
AND coord_pairs.str_pairs LIKE '%эгалик+келишик%'– эгалик ва келишикдаги сўзшакл.
```

List of derivate: давлатимни / давлатингга / давлатидан / давлатнинг / давлатларида / давлатимизни / давлатингизда / давлатнинг ...

Дастурнинг ЛТ wordform_set_coord майдонида эса аффикслар тартиби киритилди. Жумладан, wordform_set_coord (lemma+*s_ya+sh_ya+sin_ya). Бу ерда lemma – асос, s_ya – сўз ясовчи аффикс, sh_ya – шакл ҳосил қилувчи аффикс, sin_ya – синтактик шакл ҳосил қилувчи аффикс. “*” белгиси аффикснинг бирикиш тартиби қатъий талаб қилинмайдиган ҳолатни англатади. Шундан сўнг сўзшаклнинг матнда учрайдиган шакли намоён бўлади: wordform_refresh (wrд). Айнан МТнинг шу жараёни машина таржимасида муҳим ўринга эга.

МТдаги асосий бирликлар: ифода плани бирликлари (графемалар); маъно ташувчи кичик бирликлар (морфемалар); маъноли қисмлар (сўзлар); маълумотнинг коммуникатив ташувчилари (гаплар) таҳлилга тортилди.

Ўзбек тилида ёзилаётган матнлар, турли соҳаларга оид ҳужжатларни АТТ қилишга эришиш мақсадида юқорида илк қадамда эътиборга олиниши керак бўлган жиҳатлар МТ доирасида кўриб чиқилди, таҳрирловчи дастурнинг лингвистик таъминотини яратиш борасида муайян хулосалар чиқарилди.

Матнни АТТ қилиш алгоритми. Маълумки, ҳар қандай дастур алгоритмлар асосида ишлайди, алгоритмлар эса маълумотлар манбаига таянади. МТ алгоритмида қуйидаги белгилардан фойдаланилади:

U_s – лингвистик таъминотдаги сўзлар базаси, $U_s = \text{"SELECT * FROM 'U_s'"};$

Y_s – ўзбек тилидаги барча ясовчи аффикслар базаси, $Y_s = \text{"SELECT * FROM 'Y_s'"};$

S_q – ўзбек тилидаги грамматик категориялар базаси, $S_q = \text{"SELECT * FROM 'S_q'"};$

S_i – W матндан ажратиб олинган сўзшакллар, $1 \leq i \leq n$, n – W матндаги сўзшакллар сони;

S_{q_j} – S_q базадаги аффикслар, $1 \leq j \leq m$, m – S_q базадаги аффикслар сони.

T_z – сўзнинг қусурли ёзилганлигини визуал кўрсатувчи ва ёзилган хато сўзга мақбул сўзшакл вариантларини берувчи махсус функция.

Қуйида гапларни токенларга ажратган ҳолда, лемма бўйича ўзбек тилидаги сўзлар базасидан изланади, топилмаса ўзбек тилидаги барча ясама сўзлар базасидан изланади. Асос ёки ясалма базадан топилгач, унга бирикиш эҳтимолидаги аффикслар IDси бўйича ўзбек тилидаги барча аффикслар базасидан олинади. Демак, таҳлил алгоритми қуйидагича бўлади.

1. S_i даги ҳар бир сўз, U_s базасидан излансин. Топилса, кейинги қадамга, акс ҳолда 5-қадамга ўтсин.

2. S_i сўзнинг U_s базадаги ID (тартиб рақами) олинсин.

3. S_i сўзнинг ID рақамига тўғри келадиган аффикс S_{q_j} – S_q базадан излансин.

4. $S_i + S_{q_j}$ тўғри бўлса 10-қадамга ўтилсин, акс ҳолда кейинги қадамга ўтилсин.

5. S_i даги ҳар бир сўз, Y_s базасидан излансин. Топилса, кейинги қадамга, акс ҳолда 3-қадамга ўтилсин.

6. S_i сўзнинг Y_s базадаги ID (тартиб рақами) олинсин.

7. S_i сўзнинг ID рақамига тўғри келадиган аффикс S_{q_j} – S_q базадан излансин.

8. $S_i + S_{q_j}$ тўғри бўлса 10-қадамга ўтилсин, акс ҳолда кейинги қадамга ўтилсин.

9. T_z функция амалга оширилсин ва 10-қадамга ўтилсин.

10. Сўз тўғри ёзилганлиги тасдиқлансин ва кейинги S_{i+1} сўзга ўтилсин.

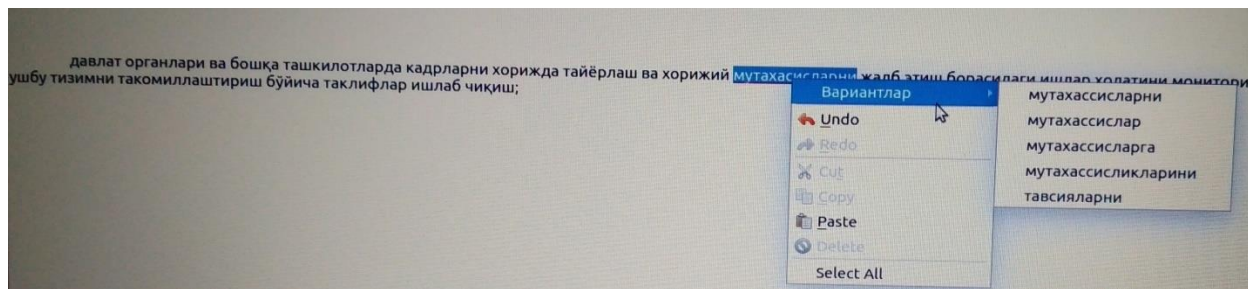
Алгоритм иш тартибини мисол орқали таҳлил қиламиз: *истеъмолчиларнинг* сўзшакли стемминг қилинади, яъни `wordform_set_coord` майдонида берилган аффикслар тартиби бўйича кесиб чиқилади: *истеъмолчи/лар/нинг*. Шундан сўнг *истеъмол* лемма имлоси List of Words базасидан текширилади, *-чи*, *-лар*, *-нинг* аффикслари S_{q_j} – S_q базасидан қидирилади, бунда *-лар*, *-нинг* шакл ясовчи ва синтактик муносабат шакли эканлиги аниқланади, *-чи* ясовчи аффиксни топиш учун Y_s базасига мурожаат қилинади. Лемма ва аффиксларнинг ЛТда мавжудлиги топилгач,

сўзнинг тўғри ёзилганлиги тасдиқланади ва кейинги сўзга ўтилади. Мисолни дастурдаги таҳлил жараёни билан бойитамиз:



давлат органлари ва бошқа ташкилотларда кадрларни хорижда тайёрлаш ва хорижий мутахассисларни жалб этиш ушбу тизимни такомиллаштириш бўйича таклифлар ишлаб чиқиш;

1-расм. Мазкур тадқиқот иши натижасида эришилган 1-амалий ютуқ – бу хато сўзларнинг тагига қизил чизиқ чизиб кўрсатилиши ҳисобланади;



2-расм. 2-катта ютуқ: хато сўзга унинг тўғри ёзилиш шакллари тақдим қилинишига эришилди.

Иккинчи бобнинг **“Омонимларни таҳлил қилиш технологияси”** параграфида омоним сўзларшакллари матн таркибида аниқлашнинг кенг қўлланиладиган усуллари ҳақида сўз юритилди. Брилл усули, Яширин Марков модели, моделлар модификацияси таҳлил қилинди, ўзбек тилидаги матнларда учрайдиган омонимларни таҳлил қилиш технологияси яратилди. Омонимларни аниқлаш бўйича ишлаб чиқилган манбалар ва формулалар асосида ўзбек адабий тили матнларидаги лексемаларни текшириш формуласи яратилди.

S – Расмий ва илмий услубдаги матнлар мажмуаси.

T_c – S матнлар мажмуаси (корпус)даги теглар сони.

w_i – бирор матндаги i – ўриндаги сўз, t_j – w сўзининг S корпусдаги теги. n – t_j теглар сони.

Матндаги w_i сўзи ва ундан кейин келган w_{i+1} сўзлари биграмма сифатида олинади.

$W_i = \{w_i, w_{i+1}\}$ – матндаги i – ўриндаги сўз ва ундан кейинги сўзлардан ташкил топган биграмма.

$Y_j = \{w(t_j), w(t_{j+1})\}$ – S корпусдаги t_j – тегли j – ўринда турган ва ундан кейинги сўзлардан ташкил топган биграмма.

Энди берилган w_i – сўзининг тўғрилиги, аффиксларнинг тўғри қўшилганлиги мавжуд S матнлар тўпламига таянган ҳолда текшириш қуйидаги формула ёрдамида амалга оширилади.

$$\prod_{j=1}^n (W_i, Y_j) \Rightarrow W_i$$

Ўзбек Миллий корпусининг яратилмаганлиги, расмий ва илмий услубдаги матнлар базаси тўла шакллантирилмаганлиги эътиборга олиниб, ўзбек тилидаги омонимшакллари таҳлил қилиш технологияси N-gramмга боғлиқ равишда яратилди. Бунинг учун ўзбек тилидаги омонимлар ЛТга ID

билан киритилди, бу жиҳат контекстда омонимларнинг грамматик шаклланишини таҳлил қилиш имкониятини беради. Таҳлил аниқлигини ошириш мақсадида омонимшаклнинг туркуми аниқланди ва унга рамзий қиймат берилди. Шундан сўнг омонимлар базаси (List of homonym) сўз бирикмалари модели (Models of CompWords)га боғланди (List of homonym=> Models of CompWords). Натижада киритилган М – омоним билан келган N – сўзга боғлиқ равишда ЛТдаги М – омонимнинг ID рақамига тўғри келувчи Q – аффиксларни олади. Яъни, M[id]+Q[id].

Мазкур усулни мисол орқали таҳлил қиламиз: $енг_1 \rightarrow N/от$ (кийимнинг қўлни қоплаб турадиган қисми), $енг_2 \rightarrow V/феъл$ (ғолиб келмоқ, устун келмоқ), $енг_3 \rightarrow V/феъл$ (ейишга ундовчи буйруқ феъли) омонимшакллар матнда синтактик модулда яратилган сўз бирикмалари моделидаги сўз туркумлари рамзларига мос равишда ўзидан кейин келган сўз (биграмма – икки белги кетма-кетлиги) билан синтактик боғланишда қабул қиладиган грамматик воситаси таҳлил қилинади. Демак, бу ерда N ва V билан боғлиқ моделларга мурожаат қилинади: $N_{com-ни} + V(енгини кесмоқ)$, $N_{com-га} + V(енгига қадамоқ)$, $Прон_{ни(дир)} + V(хаммани енгмоқ)$, $Прон_{дан(дир)} + V(мендан енгилмоқ)$, $Adj_{дан} + V(Vh)$ (кучсиздан енгилмоқ) ва ҳ.к.

Умуман, омоним-таҳлил усули муайян омоним сўзга бошқа сўзнинг маълум грамматик кўрсаткич орқали боғланиши билан тушунтирилади ва бу жараён таҳлил имконияти ҳамда таҳлилнинг аниқлигини оширади.

Мазкур бобнинг “**Дастурнинг синтактик таҳлил қилиш модули**” деб номланган бўлимида синтактик тизимлар ва синтактик боғланиш меъёрлари берилди. Маълумки, гап (жумла) миқдори чексиз ва у индивидуал тузилишига ёхуд қўлланилишга боғлиқ. Шу боис синтактик таҳлилда нисбатан кичик бирлик ҳисобланмиш сўз бирикмаларига таянилди, яъни МТ модулларини яратиш жараёнида барча лексемалар сўз туркумларига ажратилганига асосланиб, сўзларнинг боғланиш йўллари моделлаштирилди. Натижада расмий ва илмий матнларда битишув йўли билан сўзлар боғланишининг 14 модели, мослашув йўли боғланишнинг 16 модели, бошқарувли бирикмаларнинг 35 модели яратилди.

Умуман олганда, сўз бирикмалари моделлари бизга 2 мақсадда керак:

- 1) N-граммалар сўзлар бирикмасини таҳлил қилиш;
- 2) Омонимшаклларнинг лингво-таҳлил технологиясини амалга ошириш.

Ҳар бир белги ҳисобга олиниши ва белгилар тегга бирлаштирилиши ҳисобга олиниб, лингвистик таъминотда жумла қуйидагича қоидалаштирилди:

Қоида: синтактик анализаторда жумла (гап) деб, икки кўрсаткич (тиниш белгилари) орасида жойлашган матннинг қисмига айтилади. Бу кўрсаткичлар қуйидагилар: нукта / ундов белгиси / сўроқ белгиси + бўшлиқ ва нукта / ундов белгиси / сўроқ белгиси.

Схемаси:

[. / ! / ?] + [пробел] [. / ! / ?]

Мисол:

[. / ! / ?] + [пробел] *Сувни тежаб ишлатиш лозим.*

Гапларнинг синтактик тузилишини автоматик таҳлил қилиш алгоритми тавсифи берилди.

Диссертациянинг **“Сўз туркумлари модуллари”** бобида ўзбекча матнларни таҳрир ва таҳлил қилиш дастури таъминоти учун ўзбек тили сўз туркумлари модулида лексик-грамматик кодни шакллантириш мақсадида фонологик, морфонологик ва орфографик қоидалар алгоритми тузилди. Ўзбек адабий тили сўз туркумларига мансуб тубандаги кўринишлари аниқланди ва АТТ дастурининг List of Words қисмига жойлаштирилди: 3000 дан ортиқ от, 4000 дан ортиқ феъл, 1254 та сифат, 488 та равиш, 175 та олмош, 142 та тақлид сўз, 95 та ундов, 63 та модал сўз, 94 та кўмакчи, 47 та боғловчи, 36 та юклама ва 114 та ҳисоб сўзи.

Тадқиқотда исмларнинг синтактик шакл ҳосил қилувчи аффиксларни бириктирганда юзага келадиган лингвистик ҳодисалар КЛ талабларига биноан мавзулар доирасида 107 қоида шакллантирилди.

Диссертациянинг **“Лингвистик процессорда аффиксация модули”** параграфида 4 сўз туркуми (от, сифат, феъл, равиш)га хос бўлган аффиксация усули берилди.

Маълумотлар базасида аффикс қўшилаётган лексема асос ёки лемма, ҳосил бўлган лексемашакл ясалма дейилади.

Ўзбек адабий тилида сўз ясовчи аффикслар лексемага қуйидаги тартибда қўшилади: (префикс₁)+асос₂+(ясовчи аффикс₃) (*хуш₁хабар₂; қабул₂хона₃; бе₁одоб₂лик₃*)

Дастурнинг лингвистик таъминотида ўзбек адабий тилида мавжуд барча ясовчи аффикслар киритилди. Бу келажакда таҳлил дастурини мукамаллаштиришда ва янги-янги сўз ясалишларни автоматик таҳлил қилишда қулайликни яратади.

Ўзбек адабий тилида аффиксация усули билан от ясовчи 114 та аффикс, феъл ясовчи 58 та аффикс, сифат ясовчи аффиксларнинг 117 кўриниши, равиш ясовчи 48 та аффикс аниқланди.

ХУЛОСА

Ўзбекча матнларни таҳрир ва таҳлил қилиш дастурининг лингвистик таъминотини яратишда қуйидаги хулосаларга келинди:

1. Шу пайтгача амалга оширилган илмий тадқиқот ишларида автоматик таҳрир ва таҳлилга қисман эътибор берилган, холос. Мавжуд изланишларда икки тил ёхуд ўзга тил морфологияси қиёсий-типологик услубда таҳлил қилинган. Шунга қарамай, бу масала лингвистик тизим сифатида азалдан тилшуносларни қизиқтирганини кўриш мумкин.

2. Жаҳон КЛдаги матнлар билан боғлиқ тадқиқотлар, ўз навбатида, матннинг таҳрири ва таҳлили тадқиқи сари етаклайди, уларда келинадиган хулосалар ва олиннадиган натижаларнинг асосли ҳамда ишончли бўлишига

кўмаклашади. Ана шу ғояларнинг амалга ошиши компьютер лингвистикасида матнларни қайта ишлаш тадқиқотларини янги поғонага олиб чиқади ва унинг амалиётга татбиқ этилишини осонлаштиради.

3. АТТ дастурининг лингвистик таъминотини яратиш лингвистик меъёрлар ва алгоритмга эга қоидаларнинг ишлаб чиқилиши, лексикографик манбаларнинг таъминотга киритилиши, сўзшакллар ва сўзларнинг ўзаро боғланиш моделлари тузилиши билан белгилади. Мазкур жараёнда расмий ва илмий услубдаги матнлар моҳияти лингвистик таҳлил қилинмайди, балки мазкур услубдаги матнлар имлоси ва грамматикасини текширувчи система ёки ДТни яратиш тамойиллари ишлаб чиқилади.

4. Бир сўзнинг услублар доирасида ўз қўлланиш вариантыга эгаллиги ва ушбу услубларнинг меъёрларини ҳисобга олиб айтиш мумкинлиги, сўзнинг ҳамма услубдаги кўриниши ҳамда аффикслар билан келиш шаклини яратиш мураккаб вазифа. Шу боис АТТ дастурининг ЛТни яратишда илмий ва расмий услубларда қўлланиладиган сўз ва аффиксларнинг олиниши ишнинг самарадорлигини оширишга хизмат қилади.

5. Ҳар қандай дастурнинг лингвистик базасини яратишда керакли манба – “Ўзбек тилининг морфологик луғати”дир. Бундай луғатда бир лексеманинг қайси туркумга оидлиги, унинг сўзшакллари келтирилади. Бу эса дастурнинг мукамал чиқишини таъминлайди. Шу боис ўзбек тилидаги матнларга ишлов берувчи дастурларнинг лингвистик таъминотини яратишда қулайлик ва натижага тезкор эришиш мақсадида ушбу луғатнинг ишлаб чиқилиши мақсадга мувофиқдир. Морфологик луғат тилшуноснинг вақтини тежайди ва белгиланган натижага эришишни тезлаштиради.

6. Лингвистик процессорни яратишда лингвистик модулнинг ўрни ва аҳамияти ўта муҳим. КЛда модуль термини дастурий таъминотнинг муайян лингвистик жараёнини қамраган мустақил таркибий қисми сифатида қўлланилади. Лексик модулда луғат қатлами қуршаб олинса, графематик модулда рамзлар, тиниш белгилар, ҳарфий ва бошқа белгилар таҳрир қилинади, орфографик модулда имло қоидалари, морфологик модулда сўзшакллар анализи (сўзшаклдан лексемага қадар таҳлил) ва синтези (лексеманинг грамматик шаклланиши таҳлили жараёни), синтактик модулда суперсинтактик бирлик – гап ёки сўзларнинг ўзаро боғланиш ҳодисаси таҳлил қилинади.

7. Ўзбек тилининг АТТнинг дастлабки жорийланишида расмий-идоравий ва илмий услубдаги матнларни таҳрир ва таҳлил қилишга мўлжалланган лингво-системанинг мукамал ишлаши ва натижаси самарали бўлиши учун мазкур система графематик, морфологик ва синтактик таҳлил босқичларига ажратилиб, модуллаштирилиши мақсадга мувофиқ. Графематик босқичда таҳрири амалга ошириладиган бирликларнинг белгилаб олиниши таҳрир аниқлигини таъминлайди. Морфологик таҳлил принциплари от туркуми мисолида ёритиб берилиши бошқа туркум модулларини тузиш тамойилларини яратишга замин яратади. Синтактик таҳлил жараёнида гап миқдори чексиз ва у индивидуал тузилиши ёхуд қўлланилишга боғлиқ бўлгани сабабли СТда нисбатан кичик бирлик

ҳисобланмиш сўз бирикмаларига таянилган маъқул. Яъни, МТ модулларини яратиш жараёнида барча лексемалар сўз туркумларига ажратилганига асосланиб, сўзларнинг боғланиш йўллари моделлаштирилади ва “сўз бирикмалари категорияси” яратилади.

8. Сўзшаклларни юзага келтиришда орфографик, фонологик, морфонологик ҳодисалар рўй беради. Лингвистик модулни яратишда ҳозирги ўзбек адабий тилининг лингвистик меъёрларига матн таҳрири ва таҳлилида асосий лисоний ҳодиса сифатида қаралади. Анъанавий тилшунослик меъёрлари асосида КЛда ўзбек тили сўз туркумлари модулида лексик-грамматик кодни шакллантириш мақсадида фонологик, морфонологик ва орфографик қоидалар алгоритмини тузиш; синтактик ва семантик кодни шакллантирувчи сўз бирикмаларининг лингвистик моделини яратиш грамматик қоидаларни бериш, уларнинг алгоритмини яратиш устувор вазифа ҳисобланади. Мазкур масала нафақат лингво-таҳлил жараёнининг сифатли олиб борилишини таъминлайди, балки ўзбек тилини ўқитиш учун ҳам долзарблигини кўрсатади.

9. Ўзбек тилидаги ясовчи қўшимчаларни статистик жиҳатдан ҳисоблаш ва уларнинг лексик базаси ва *лемма+ясовчи аффикс модели*ни яратиш лингвистик таъминотнинг муҳим таркибий қисми ҳисобланади. Компьютер учун аффиксларни сўз ясовчи ва шакл ясовчи аффикслар тарзида алоҳида базаларга ажратиш, моделларини аниқлаш таҳлил жараёнида муҳим.

10. Омоним шакллар ҳамиша тилшунослик марказидаги ҳодисалардан бири сифатида қаралган. Шу боис матнда учрайдиган омошакллар рус ва хориж тилшунослигида алоҳида тадқиқ қилинади. Шу кунга қадар ишлаб чиқилган омоним сўзшаклларни матндан аниқлаш ёхуд уларни бартараф этиш усулларидаан Брилл усули, яширин Марков модели, N-gramm ва бошқа моделлар модефикацияси ўрганилиб, ўзбекча матнларда омоним шаклларни таҳлил қилиш технологияси татбиқ қилинди. Натижада омонимшаклларни матндан аниқлаб уларни таҳлил қилиш асоси яратилди.

11. Фразеологик бирикмалар нутқий эмас, балки лисоний ҳодиса эканлиги, улардаги грамматик боғланиш турғун бўлганлиги боис уларнинг категориал моделларини яратиш мураккаб масала ҳисобланади. Шу боис туғун бирикмалар таркибидаги сўзларнинг боғланиш модели эмас, балки грамматик шаклини ўзгартириш ўринларини белгилаш ва лингвистик базага киритиш автоматик таҳлилнинг аниқлигини таъминлайди.

12. Лингво-анализ дастури нафақат таҳрир ва таҳлил жараёнини тезлаштирувчи восита, балки ўзбекча матнларни беқусур ёзиш кўникмасини шакллантирувчи инновацион система сифатида муҳим аҳамиятга эга ҳисобланади.

13. Ҳозирга қадар жамланган лингвистик қоидаларнинг кенг кўламли тадқиқи келажакда ўзбек лингвистик корпусининг яратилишига муҳим ҳисса бўлиб қўшилади.

**ТАШКЕНТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
УЗБЕКСКОГО ЯЗЫКА И ЛИТЕРАТУРЫ**

АБЖАЛОВА МАНЗУРА АБДУРАШЕТОВНА

**ЛИНГВИСТИЧЕСКИЕ МОДУЛИ ДЛЯ ПРОГРАММЫ
РЕДАКТИРОВАНИЯ И АНАЛИЗА ТЕКСТОВ НА УЗБЕКСКОМ
ЯЗЫКЕ (ДЛЯ ПРОГРАММЫ АНАЛИЗА ОФИЦИАЛЬНЫХ И
НАУЧНЫХ ТЕКСТОВ)**

10.00.11 – Теория языка. Прикладная и компьютерная лингвистика

**АВТОРЕФЕРАТ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ ДОКТОРА
ФИЛОСОФСКИХ НАУК ПО ФИЛОЛОГИИ (PhD)**

Тема диссертации на соискание ученой степени Доктора философии (PhD) зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за № B2019.2.PhD/Fil.320

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

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

Научный руководитель: **Дадабоев Хамидулла Арипович**
Доктор филологических наук, профессор

Официальные оппоненты: **Улуков Носиржон Мухаммадалиевич**
Доктор филологических наук, профессор

Пардаев Азамат Бахронович
Доктор филологических наук

Ведущая организация: **Бухарский государственный университет**

Защита диссертации проводилась на заседании Научного совета № DSc.27.06.2017.Fil.05.02 по присуждению ученых степеней при Ферганском Государственном Университете «__» _____ час __ мин. (Адрес: 100151, г. Фергана, ул. Мураббийлар, 19. Тел: (993873) 244-44-29; факс: (+99873) 244-66-03; e-mail: fardu_info@umail.uz).

С текстом диссертации можно ознакомиться на сайте Информационно-ресурсного центра Ферганского Государственного Университета (регистрационный № _____). (Адрес: 100151, г. Фергана, ул. Мураббийлар, 19. Тел: (99873) 244-71-28).

Автореферат диссертации разослан « "_____ " _____ 2019 г.

(протокол реестра № _____ от "_____ " _____ 2019 г.

М.Х.Хакимов

председатель Научного совета по присуждению ученых степеней, д.ф.н., профессор

М.Т.Зокиров

научный секретарь Научного совета по присуждению ученых степеней, к.ф.н., доцент

С.Муминов

председатель Научного семинара при научном совете, д.ф.н., профессор

ВВЕДЕНИЕ (аннотация к диссертации доктора философских наук (PhD))

Актуальность и необходимость темы диссертации. С 50-х годов XX века в мировой лингвистике стала изучаться проблема редактирования текста с помощью электронных вычислительных машин и в настоящее время проводятся значительные исследования для создания лингвистического обеспечения для редактирования и анализа современных информационных технологий, программного обеспечения и грамматических правил естественного языка. Известно, что технический прогресс на этой основе привел к созданию систем количественного редактирования и анализа на ведущих литературных языках мира.

В современной мировой лингвистике автоматизированного анализа и редактирования (ААР) разработаны лингвистические средства для автоматического редактирования. Они являются не только средствами быстрого и экономного редактирования и анализа, но в тоже время способствуют улучшению качества машинного перевода. Поскольку в веке глобального развития научной техники необходимыми факторами являются работа с большим объёмом текста и получение эффективных результатов в кратчайшие сроки. Следовательно, перед мировой лингвистикой ставятся задачи лингвистического моделирования языка, формирование леммы языка, создание алгоритма редактирования и анализа текста, а также формирование лингвистического модуля программы лингвоанализа.

В узбекской лингвистике было проведено несколько исследований по отбору, анализу и предобработке текстов с помощью информационных технологий, но существует потребность в эффективной международной практике в этой области исследований и лингвистическом обеспечении текстового редактирования и анализа. Действительно, «...наша задача – создать эффективные механизмы для реализации результатов научных исследований, устойчивого развития науки и инновационной деятельности».²⁵ Именно среди этих задач стоит создание лингвистических модулей программы автоматически редактирование и анализа текстов с помощью компьютерных средств. Это свидетельствует о том, что тема данного исследования чрезвычайно актуальна для узбекской лингвистики.

Исследования, проведенные в рамках диссертации, способствуют осуществлению задач, определенных Указом № УП-4947 Президента Республики Узбекистан “О стратегии действий по дальнейшему развитию Республики Узбекистан” от 7 февраля 2017 года, Постановлением № ПП-2789 “О мерах по дальнейшему совершенствованию деятельности академии наук, организации, управления и финансирования научно-исследовательской деятельности” от 17 февраля 2017 года, Указом УП-4797 “Об организации деятельности Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои” от 13 мая 2016 года, Постановлением № ПП-3775 “О дополнительных мерах по повышению

²⁵ Мирзиёев Ш. Миллий тараққиёт йўлимизни қатъият билан давом эттириб, янги босқичга кўтарамиз. I жилд. – Тошкент: Ўзбекистон, 2017. – Б. 168-174.

качества образования в высших образовательных учреждениях и обеспечению их активного участия в осуществляемых в стране широкомасштабных реформах” от 5 июня 2018 года а также другими нормативно-правовыми документами, регулирующими данную деятельность.

Соответствие исследования приоритетным направлениям развития науки и технологии в Республике. Настоящее исследование осуществлено в соответствии с приоритетными направлениями I. «Формирование системы инновационных идей и способов их осуществления при социальном, правовом, экономическом, культурном и духовно-просветительском развитии информатизированного общества и демократического государства».

Степень изученности проблемы. Вопрос текстов ААР в области компьютерной лингвистики за рубежом изучался многими учеными.²⁶ Дж.Петерсон упоминал о программных обеспечениях, которые способны проверять и исправлять ошибки текстов. Изучение текстового анализа в лингвистическом аспекте связано с именем Д.Юрафского. Исследования ученого по автоматизированной обработке текстов и автоматическому анализу²⁷ стали ключевым фактором и дали дальнейший акцент на будущие исследования.

Программы АР были разработаны для автоматического редактирования русского, английского, немецкого, французского и итальянского текстов. Даже морфологические и синтаксические анализаторы были созданы для таких текстов, как тюркский, туркменский, казахский, киргизский и азербайджанский, а в систему MS WORD включены пакеты автоматической модификации.²⁸ Такие программы редактирования в основном предназначены для проверки правописания. В области автоматического анализа русского языка провели эффективное исследование ученые такие как, Р.Г.Котов, В.Э.Берзон, В.Г.Бритвин, И.А.Мельчук, А.Б.Кузнецов, Л.И.Беляева В.А.Чижаковский, Г.Г.Белоногов, И.С.Дуганова, Л.Л.Иомдин,

²⁶Peterson, James. Computer Programs for Detecting and Correcting Spelling Errors. 1980.; Do Prado H.A. Emerging Technologies of Text Mining: Techniques and Applications / Ed. by H. A. Do Prado, E. Ferneda. – Idea Group Reference, 2007. – 358 p.; Earnest, Les. The First Three Spelling Checkers. Stanford University. Archived from the original (PDF) on 22 October 2012.; Earnest, Les. Visible Legacies for Y3K (PDF). Archived from the original.; de Amorim, R.C.; Zampieri, M. Effective Spell Checking Methods Using Clustering Algorithms. Proceedings of Recent Advances in Natural Language Processing. Hissar, Bulgaria. 2013. – P. 172-178.; Zampieri, M.; de Amorim, R.C. Between Sound and Spelling: Combining Phonetics and Clustering Algorithms to Improve Target Word Recovery. Proceedings of the 9th International Conference on Natural Language Processing (PolTAL). Lecture Notes in Computer Science (LNCS). Springer. 2014. – P. 438-449.

²⁷ Jurafsky D., Martin J. Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. – New Jersey: PrenticeHall, 2000. – 927 б.

²⁸ Орехов Б.В., Слободян Е.А. Проблемы автоматической морфологии агглютинативных языков и парсер башкирского языка [Текст] // Информационные технологии и письменное наследие: материалы международной научной конференции (Уфа, 28–31 октября 2010 г.) / отв. ред. В. А. Баранов. – Уфа. Ижевск: Вагант, 2010. – С. 167-171.; <http://lcpb.bashedu.ru/index.php?go>. Орехов Б., Галлямов А. Башкирский морфологический анализатор (Электронный ресурс); <http://my.ict.nsc.ru/~salerat/kaz/>. Программа генерации словоформ казахского языка; <http://tamgasoft.kg/morfo/ru/>. Морфология киргизского языка; Сулейманов, Д. Ш. Синтаксический анализатор предложений татарского языка / Д. Ш. Сулейманов, А. Р. Гатиатуллин // Труды Математического центра имени Н. И. Лобачевского. – Казань, 1999. – Т. 4. – С. 111-126.

Э.К.Лавошникова, А.В.Сокирко, И.М.Ножов, Е.И.Большакова, Д.В.Клышинский, Н.Н.Леонтьева²⁹ и др.

Особое внимание уделяется исследованию в области лингвистического анализа текстов тюркских языков статистическими методами, ученых, таких как Р.Г.Пиотровский, М.А.Махмудов, А.А.Пиотровская, Т.Садиков³⁰. Ученые Х.Арзикулов, С.Ризаев и С.Мухамедов, провели предварительные исследования по компьютерной лингвистике (КЛ) на узбекских материалах.³¹ В основном они выполняли теоретическую и практическую работу по статистическому анализу КЛ.

В начале XXI века были проведены ряд исследований на узбекском и английском языках, а также были опубликованы десятки научных статей, учебников³² и книг³³, успешно защищены диссертационные работы, кандидатские диссертации, посвященные проблемам компьютерного обеспечения узбекского языкознания, под руководством А.Пуллатова, внесшего значительный вклад в развитие узбекской КЛ. Особенно были проведены много работ, посвященных формированию класса глаголов, разработке их модели и созданию их с помощью глаголов.³⁴ Лексический и

²⁹Котов Р.Г. Лингвистические аспекты автоматизированных систем управления. – Москва: Наука, 1977. Он же. Прикладная лингвистика и информационная технология. – Москва: Наука, 1987.; Котов Р.Г., Якунин Б.В. Язык информационных систем. – Москва: 1989.; Он же. Оптимизация речевого воздействия. – Москва: Наука, 1990.; Берзон В.Е. и др. О разработки подсистемы редактирования выходного текста в рамках системы МП // Проблемы внутренней динамики речевых норм. – Минск, 1992, – С. 169-174; Бритвин В.Г. Прикладное моделирование синтагматической семантики научно-технического текста (на примере автоматического индексирования): Дисс. канд. филол.наук. – Москва: МГУ, 1983.; Мельчук И.А. Порядок слов при автоматическом синтезе русского слова (предварительные сообщения) / Научно-техническая информация. 1985, – №12. – С. 12-36.; Беляева Л.И., Чижаковский В.А. Тезаурус в системах автоматической переработки текста. – Кишинев, 1983.; Белоногов Г.Г., Котов Р.Г. Автоматизированные информационно-поисковые системы. – Москва, 1968; Белоногов Г.Г., Дуганова И.С., Кузнецов А.Б. Экспериментальная система автоматизированного обнаружения и исправления ошибок в тексте / НТИ. Серия 2, 1984, – №3. – С. 20-25.; Иомдин Л.Л. Автоматическая обработка текста на естественном языке: модель согласования. – Москва: Наука, 1990. – 180 с.; Лавошникова Э.К. О «подводных камнях» в компьютерных системах проверки правописания // Вестник Московского университета. Серия 9. Филология. 2002, – №6. – С.151-163.; Сокирко А.В. Семантические словари в автоматической обработке текста: По материалам системы ДИАЛИНГ: Дисс. ... канд.тех.наук. – Москва, 2001. – 120 с.; Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы) сегментации русского предложения: Автореф. дисс. ...канд. филол.наук. – Москва, 2003.; Большакова Е.И., Клышинский Э.С. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб.пособие. – Москва: МИЭМ. 2011. – 272 с.; Леонтьева Н.Н. Автоматическое понимание текста: системы, модели, ресурсы: учеб.пособие. – Москва: Академия. 2006. – 304 с.

³⁰Махмудов М.А., Пиотровская А.А., Садыков Т. Система системного анализа и синтеза тюркской словоформы / Переработка текста методами инженерной лингвистики. – Минск, 1982.; Пиотровский Р.Г. и др. Методы автоматического анализа и синтеза текста. – Минск: Высшая школа, 1985.

³¹Арзикулов Х.А., Пиотровская К.Р. Информатика и переработка текста средствами вычислительной техники (учебное пособие). – Самарканд, 1986.; Ризаев С. Ўзбек тилининг лингвостатистик тадқиқи: Филол. фан. док. ...дисс. автореф. – Тошкент, 2008. – 50 б.; Шу муалл. Ўзбек тилшунослигида лингвостатистика муаммолари (монография). Тошкент: Фан. 2005.; Мухамедов С.А. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дис. ...канд.филол.наук. – Тошкент, 1980. – 25 с.

³²Мухамедова С. Компьютер лингвистикаси (методик қўлланма). – Тошкент, 2007.; Пуллатов А., Мухамедова С. Компьютер лингвистикаси (ўқув қўлланма). – Тошкент, 2008. – 98 с.

³³Пуллатов А. Компьютер лингвистикаси. – Тошкент: Akadernashr, 2011. – 520 б.; Rahimov A. Kompyuter lingvistikasi asoslari. – Toshkent: Akadernashr. 2011. – 160 б.; Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.

³⁴Пуллатов А.К., Алиходжаев Б., Джураева Н. Разработка программы компьютерного анализа и синтеза глаголов узбекского языка // ЎзМУ хабарлари. – Тошкент, 2002. – №2. – С. 17-19.; Po'latov A.Q., Mo'minova

синтаксический анализ узбекского и русского языков проводился в последние годы под руководством М.А.Хакимова для системы «Переводчик L-MX».³⁵ Н.Абдурахмоновой, Ш.Хамраевой и Д.Уринбаевой³⁶ были даны конкретные комментарии по анализу текстов, что подчеркивает важность разработки лингвистических модулей в создании лингвистических программ.

В результате этих исследований были собраны теоретические знания и конкретный опыт для текстов ААР. Основываясь на этих исследованиях, мы попытались создать автоматизированную редакционную программу в узбекской КЛ, что является первой инициативой в этом исследовании. Лингвистические модули автоматического редактирования программ создают совершенное программное обеспечение, которое помогает готовить грамотные литературные тексты на узбекском языке.

Связь темы диссертации с научно-исследовательскими работами высшего учебного заведения, где была выполнена диссертация. Диссертация была осуществлена согласно плану научно-исследовательских работ Ташкентского государственного университета узбекского языка и литературы им. Алишера Навои «Изучение социального, исторического и современного языкового развития».

Целью исследования является создание лингвистических модулей программы графематического редактирования и морфологического, синтаксического анализа официальных и научных текстов узбекского литературного языка, а также формирование лингвистического обеспечения узбекского языка.

Задачи исследования:

выявление и классификация содержания и сущности передовых технологий, использующих процессы компьютерного редактирования и машинного анализа;

создание алгоритма фонологических, морфологических и орфографических правил формирования лексико-грамматического кода узбекского языка; создание лингвистической модели формул синтаксического и семантического кода;

изучение этапов автоматического лингвистического анализа и разработка на этой основе модули этапов программирования, а также выявление и решение проблем лингвистического редактирования и анализа;

Т., Po'latova I.O. Duniyoviy o'zbek tili (O'zbek tilida fe'lining shakllari va ularning rus, ingliz tillarida berilishi). – Toshkent: Universitet, 2003. – 404 b.; Мухамедова С. Ўзбек тилидаги ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. – Тошкент, 2006. – 80б.

³⁵ Хакимов М.Х. Расширяемый входной язык математического моделирования естественного языка для много-язычной ситуации машинного перевода // ЎзМУ хабарлари, 2009. – № 1. – С.75-80.

³⁶ Абдурахмонова Н. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти: Филол.фан.бўйича фалсафа доктори (PhD)...дис. афтореф. – Тошкент, 2018. – 47 б.; Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол.фан.бўйича фалсафа доктори (PhD)...дис. афтореф. – Қарши, 2018. – 53 б.; Ўринбоева Д. Халқ оғзаки ижоди: жанрий-лисоний ва лингвостатистик тадқиқ муаммолари: Филол. фан. бўйича докт. (DSc) ... дисс. автореф. – Самарқанд, 2019. – 74 б.

создание оптимальных способов разработки языковой поддержки программного обеспечения и создание программного языкового алгоритма с использованием модулей.

В качестве объекта исследования были выбраны тексты, написанные в официальных и научных стилях на узбекском литературном языке.

Предметом исследования является разработка основ языкового обеспечения автоматизированного редактирования узбекских текстов и программного обеспечения для анализа.

Методы исследования. Тема характеризовалась описательным, компонентным анализом, статистическим, сравнительно-типологическим, моделированием, древообразным синтаксическим анализом, методами аналогии.

Научная новизна исследования:

впервые в узбекской компьютерной лингвистике была создана теоретическая основа для лингвистического обеспечения системы ААР и раскрыта лингвистическая, практическая и образовательная ценность системы ААР;

при создании автоматического редактирования и лингвистического обеспечения анализа разработаны принципы разделения синтагматического и методического знаков и морфемной валентности, обоснованы модули и значения модуляции;

обеспечение грамматических категорий узбекского языка по научным и официальным текстам, сгруппирование по частям речи, соединение в составе фразы, сформирования модели соединения с помощью управления и согласования;

разработаны модули технологии автоматического редактирования, анализа и графематического редактирования, обеспечивающий совершенствование программы, морфологического и синтаксического анализа.

Практические результаты исследования:

собраны фразеологические единицы и слова омонимы, имеющиеся в узбекском языке, сформирована база, определены факторы отличия их, а также изобразительные выражения и варианты слова, статистика словообразовательных аффиксов узбекского языка;

создана технология анализа омонимов, формула проверки лексем в текстах;

составлен алгоритм морфологического анализа и принципов создания лингвистической базы для программного обеспечения, лексикографических источников лингвистического обеспечения автоматического редактирования и анализа;

разработана программа автоматического редактирования и анализа узбекских текстов.

Достоверность результатов исследования основывается на теоретических выводах и определяется сочетанием выводов по лингвистике и компьютерной лингвистике, тематических публикациях на других языках,

актуальности практической значимости работы, а также на «Толковом словаре узбекского языка», «Толковом словаре узбекских омонимов» и «Современный литературный узбекский язык».

Научная и практическая значимость результатов исследования. Изучение лингвистических вопросов в области ААР служит для анализа методологических теорий по конкретным аспектам КЛ, расширения сферы исследований, создания модулей ЛО и обучения узбекскому языку в образовательных учреждениях. Результаты исследований и научно-теоретические выводы могут быть использованы на курсах и семинарах для высших учебных заведений и среднего профессионального образования. Исследование имеет научное значение при создании монографий, учебников и учебных пособий, связанных с ААР.

В результате использования ЛМ и их алгоритмов, разработанных в этом исследовании, обеспечивается создание программы для редактирования официальных и научных текстов на узбекском языке. А также служит фундаментальным ресурсом для создания лингвистической базы данных программного обеспечения для обработки текста. Он предназначен для курсов по компьютерной лингвистике, прагмалингвистике, когнитивной лингвистике, а также организации семинаров и практикумов для практических занятий. Эти аспекты определяют практическую значимость диссертации.

Применение результатов исследования. Результаты составления официальных и научных текстовых редакционных изданий, полученные в диссертационном исследовании, были представлены в следующих областях:

Диссертация была использована в «Создание многоязычных словарей электронной и банковско-финансовой терминологии, связанной с узбекским языком и его переводом» в результате перевода лексики и статистики и формального подхода к узбекскому языку. (Справка № 89-03-915 Министерства Высшего и среднего специального образования от 5 марта 2019г.). Были усовершенствованы многоязычные словари на основе научных результатов;

Анализ и выводы по синтаксическому анализу с использованием программ ААР и моделей словосочетаний были использованы в Приложении к проекту государственных грантов на тему «Создание аудиокниг по предметам лингвистики для учащихся-инвалидов» (Справка № 89-03-915 Министерства Высшего и среднего специального образования от 5 марта 2019г.) В результате была разработана система лабораторных заданий для раздела синтаксиса «Современного узбекского литературного языка».

Материалы диссертации были использованы в эфире телеканала «Навои» для программы «Дахлдорлик туйғуси» и радиоканала «Узбекистан» для программы «Тил – миллат кўзгуси», «Таълим ва ислоҳот». (Справка № 01-02/57 Навоийского областного телерадиокомпании от 11 марта 2019 года). В результате был обеспечен научный потенциал вещания и радиовещания.

Апробация результатов исследования. Результаты исследования были апробированы на 6 международном, 7 национальных и международных научных конференциях.

Публикация результатов исследования. По теме диссертации опубликованы 31 научных работ. В том числе, 1 научное свидетельство, в научных изданиях, рекомендованных Высшей аттестационной комиссией Республики Узбекистан для публикации основных результатов докторских диссертаций, опубликованы 5 статьи, 3 из них в республиканских изданиях и 2 в зарубежном издании.

Содержание и объем диссертации. Диссертация состоит из введения, 3-х глав, заключения и списка использованной литературы. Общий объем составляет 164 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ИССЛЕДОВАНИЯ

Во **вводной части** диссертации обоснованы актуальность темы диссертации и ее необходимость, описаны цели и задачи исследования, его объект и предмет, соответствие работы приоритетным направлениям развития науки и технологии в республике, обоснована научная новизна, описано практическое значение, освещены научная и практическая значимость полученных результатов, приведены сведения о практическом использовании результатов исследования, опубликованных работах, и описано содержание исследования.

В первой главе диссертации «**Анализ и редакция текста в парадигме когнитивных исследований**» был проведен ряд исследований, направленных на создание компьютерных моделей для понимания естественного языка в этой области и в области когнитивного образования.

В русской компьютерной лингвистике проведены эффективные исследования по разработке программного обеспечения для анализа текста. В частности, ряд работ посвящен общим проблемам автоматизированного анализа в области КЛ, принципам моделирования³⁷, морфологического анализа³⁸, синтаксического³⁹ и семантического⁴⁰ анализа, аналитических единиц, теории автоматического понимания текстов⁴¹ на естественном языке,

³⁷Ягунова Е.В. Основы теоретической, вычислительной и экспериментальной лингвистики / Е.В.Ягунова // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие / Е. И. Большакова [и др.]. – Москва: МИЭМ. 2011. – 272 с. – С. 7-89.

³⁸Клышинский, Э. С. Начальные этапы анализа текста / Э. С. Клышинский // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие / Е. И. Большакова [и др.]. – Москва: МИЭМ. 2011. – 272 с. – С. 106-140.; Марчук, Ю. С. Компьютерная лингвистика / Ю. С. Марчук. – М., 2007. – С. 60-70.

³⁹Компьютерный синтаксический анализ: описание моделей и направлений разработок / Г. Д. Карпова [и др.] // Итоги науки и техники. Серия «Вычислительные науки». Т. 6. – Москва, 1991. – 243 с.; Иорданская, Л. Н. Автоматический синтаксический анализ. Т. 2. Межсегментный синтаксический анализ / Л. Н. Иорданская. – Новосибирск, 1967. – 231с.; Мельчук, И. А. Автоматический синтаксический анализ. Том 1. Общие принципы. Внутрисегментный синтаксический анализ. / И. А. Мельчук. – Новосибирск, 1964. – 360 с.

⁴⁰Апресян, Ю.Д. Исследования по семантике и лексикографии. Т. I: Парадигматика. – М.: Языки славянских культур, 2009.; Сокирко А.В. Семантические словари в автоматической обработке текста: По материалам системы ДИАЛИНГ: Дисс. ... канд.тех. наук. – Москва, 2001. – 120 с.

⁴¹Леонтьева Н.Н. К теории автоматического понимания естественных текстов. – Москва: МГУ, 2000.

этапам анализа текста, морфологического анализа и синтеза. Однако процесс автоматической трансформации и анализа не был изучен в узбекской КЛ.

История создания программирования для редактирования и анализа текстов на английском языке, которая широко распространена сегодня, началась американским ученым компьютерных технологий Лестером Дональдсом Эрнестом. В феврале 1971 года Л.Ральф Горин (SPELL (орфография) для Desk PDP-10), аспирант в лаборатории искусственного интеллекта Стэнфордского университета под руководством Эрнеста, создает практическую программу, а не первую проверку орфографии для текстов на английском языке.⁴² Для более быстрой работы SPELL Горин написал программу на языке программирования ассемблера. Также в диссертации представлены лингвистические модули компании “Информатик”, создающей систему ОРФО, выполняющей автоматический грамматический анализ текстов на русском языке.

Тексты на узбекском языке (первый шаг) Microsoft Office Word было разработано для автоматизированного и аналитического редактирования, а также для введения лингвистического руководства с целью предоставления лингвистических материалов целевой аудитории. Их практическое значение дано в виде краткого обзора. Также были сделаны комментарии по терминологии, связанные с этим случаем в мировой компьютерной лингвистике, поскольку она впервые была теоретически изучена в узбекской лингвистике для автоматического редактирования и анализа текстов.

Суть термина модуль, который активно используется в зарубежной компьютерной лингвистике, охватывает область применения в других областях. Лингвистический модуль является неотъемлемой частью лингвистического программирования, которое является частью определенного процесса лингвистического программирования.⁴³ Лингвистическое снабжение, другими словами, лингвистический процессор, представляет собой совокупность этих компонентов. В лексическом модуле охватывается словари, графический модуль содержит символы, знаки препинания, алфавит и другие знаки, модуль орфографии включает правила правописания, морфологический модуль анализ формы слова (анализ от формы слова до лексемы) и синтез (анализ грамматической формулировки лексемы), в синтаксическом модуле суперсинтаксическая единица – анализируется взаимодействием слов.

В связи с тем, что вопрос о реформирование нового алфавита, которая основывается на латыни стоит в повестке дня, в этом исследовании правила правописания и лексические источники были основаны на кириллице.

Лингвистический словарь и грамматические правила узбекского литературного языка необходимо будет использовать для создания лингвистического процессора для автоматического редактирования и анализа

⁴²Peterson, James. Computer Programs for Detecting and Correcting Spelling Errors.1980. https://en.wikipedia.org/wiki/Spell_checker

⁴³ Демьянков В.З. Когнитивная лингвистика как разновидность интерпретирующего подхода // Вопросы языкознания. – Москва, 1994. – № 4. – С. 22.

текстов на узбекском языке. В процессе этого исследования было создано 105 перифраз (изобразительное выражение), 1393 фразем в форме грамматических форм языкового снабжения, где была определена роль грамматических форм. А также, в целях выполнения омоним форм в тексте сконцентрировано 1378 омоним форм.

В своем исследовании Е.И.Большакова подчеркивает, что лингвистический процессор основан на лингвистической модели данных, то есть лингвистических ресурсах ПО: компьютерные словари, грамматика языка, словарный запас, тезаурус и онтология, набор текстов и корпус.⁴⁴

По завершении компонента лингвистической поддержки ААР можно разделить на стабильные и динамические (переменные) типы. Лингвистическая поддержка постоянной программы создается исключительно экспертом, а ее неологизмы, новые лингвистические модели и языковые разработки включены в следующие версии программы. Задача динамической редакционной программы – постоянно обогащать лексикографическое программное обеспечение новыми словами. В большинстве случаев эта задача решается пользователем, то есть из-за определенного неологизма, используемого в тексте, новый термин может быть изменен и впоследствии применен по-разному, в результате включения в требуемый модуль поставки в соответствии с конкретными запросами. Это служит для разработки текста редакционной программы, потому что язык обеспечивает постоянный прогресс, чтобы отслеживать появление новых значений, неологизмов, архаизмов и историзмов. Е.Большакова рассматривает⁴⁵ эту возможность как своего рода морфологическое словарное обогащение и делит ее на два типа: 1) словарное обогащение (статический ПО); 2) без обогащения (динамический ПО).

Написание программы ААР в качестве примера инновационного образования показало, что оно обладает следующими возможностями:

- 1) быстро редактирует различные тексты;
- 2) улучшает навыки писания пользователя;
- 3) помогает пользователю самостоятельно и грамотно писать тексты;
- 4) является лингвистическим источником в изучении лексических и грамматических принципов узбекского языка.

Во второй части диссертации «**Модули автоматической системы редактирования и анализа текста**» были проанализированы тексты и принципы создания следующих языковых модулей: модуль редактирования графематики – это этап редактирования различных символов и отдельно используемых букв в тексте;⁴⁶ модуль морфологического анализа – от слова до леммы (лексема в словаре); синтаксический анализ – грамматическое связывание слов в тексте.

⁴⁴Большакова Е.И.и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие. – Москва: МИЭМ, 2011. – С. 97-99.

⁴⁵Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие. – Москва: МИЭМ, 2011. – С. 112-122.

⁴⁶ Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы: учебное пособие. – Москва, 2006. – С.50.

Желательно использовать технологию *токенизация* для перевода потока элементов, используемых в тексте на этапе графического редактирования (ГР), в слово в слово и в различные символы, для подготовки текста к процессу морфологического анализа и вставки текста в корпус. Потому что токенизация – это процесс разделения всех единиц в электронном тексте. Это можно проиллюстрировать на следующем примере:

[‘Ўзбекистон’, ‘Республикаси’, ‘Президентининг’, ‘2005’, ‘йил’, ‘15’, ‘июндаги’, ‘«’, ‘Шарқ’, ‘тароналари’, ‘»’, ‘халқаро’, ‘муסיқа’, ‘фестивалига’, ‘тайёргарлик’, ‘кўриш’, ‘ва’, ‘уни’, ‘ўтказиш’, ‘тўғрисида’, ‘»’, ‘ги’, ‘ПҚ’, ‘-’, ‘103’, ‘-’, ‘сонли’, ‘қарори’, ‘ўз’, ‘кучини’, ‘йўқотган’, ‘деб’, ‘ҳисоблансин’, ‘.’]

Здесь квадратная скобка показывает начало и конец текста. Элемент, слияния элемента черточки служит для разделения символов, разделенных запятыми. В этом примере часть токенизации «*ги*» обрезается, потому что она не имеет никакого значения. После этого процесса можно комбинировать элементы, вычислять их, группировать или удалять слова, которые являются слишком сложными или не имеют смысла.

При создании модулей ЛГ этапа программы лингвистического редактирования выявляются типы текстовых элементов.⁴⁷ Эти типы включают следующие единицы: ЛекЕ – лексическая единица, лексема букв алфавита, принадлежащих одной надписи (*valida, buyruq, yurak, hayot, oldin*); ЛекИ – лексема иностранного языка (*принтер, бюро*); ЦЦ – Цифровая целостность (*1986, 18/04/2012, 5.05, 19,25*); БЦИ – буквенно-цифровой интеграл (*Боинг – 767, СУ – 27, “Келажак овози – 2017”, “Йил аёли – 2019”*); АББР – аббревиатура (*БМТ, МДХ, ЎзР, ДХХ, ЎЗМУ*); СЕ – сокращенные единицы (*ва ҳ.к., ва бошқ., м-н.*).

История морфологического анализа (МА) и его значение. От компьютерной лингвистики к 60-70м годам прошлого столетия все исследования машинной морфологии начались с создания машинного словаря.⁴⁸

Морфологический анализ в основном относится к орфографическим ошибкам и фокусируется на словоформах. Для достижения этой цели необходимы следующие действия: 1) ввести словарь в компьютер, то есть электронный словарь; 2) отделение слов литературного языка в словаре; 3) методическая (научная, художественная, публицистическая, официальная) группировка выделенных слов; 4) деление литературных слов на категории; 5) группировки слов по группам (например, группирование слов в существительное, названия растений, географические названия и т.д.); 6) создание комбинации аффиксов сгруппированных слов; в котором комбинация аффиксов практически используется в литературном языке. Последовательность аффиксов приведена в последовательности комбинаций.

⁴⁷Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы: учебное пособие. – Москва: Академия, 2006. – С.58.

⁴⁸Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы): Дис. канд. филол.наук. – Москва, 2003. – С. 54.

Например, множественная форма существительных собственность + владение + категория и т.д. Порядок грамматики в структуре слова имеет определенную законность и порядок, а последовательность в расположении грамматического инструмента зависит от его значения и грамматического характера: *новое языковое значение – это первое (1), второе, которое влияет на языковое значение (2), и третье (3), которое не влияет на языковое значение: S+1+2+3.*

Порядок синтаксической формы в именах: собственный аффикс + падежный аффикс: *мартабасини, ўқиганимни*; в глаголах: подтверждение + время / склонность + лицо + числительное: *ўтмаяпсан*. Аффикс-частица добавляется после всех типов аффиксов: *давлатимизними, лойиҳамгами, ўқимаяпсанми*.

7) комбинация созданных аффиксов прилагается к словам; 8) вышеуказанные работы, будут выполнены лингвистами и включены в программное обеспечение программистами. Создана технология морфологического анализа в мире компьютерной лингвистики, и в настоящее время она стала технологией стеммирования и лемматизации. Желательно использовать стеминговую технологию для использования агглютинативных текстов в АРР.

Известно, притяжательные аффиксы и слова с характеристиками изменения объединяются под именами. Они включают существительное, прилагательное, числительное, местоимение, глаголы. В этом контексте лингвистические и другие грамматические категории ассоциируются со словами, которые относятся к названиям жанров и были составлены в рамках морфологического синтеза создания программы ЛО. Для этого в Access была создана база данных формул, которая создает языковую форму и синтаксическую форму (gr_form). Мультиязычные слова или семантические слова (в основном безымянные существительные, абстрактные существительные) были объединены в sol_CorrNounNumber.

Специальная секция программирования (left_GenIsmForm) была создана для грамматических образований узбекских имен. При этом 1) грамматические формы были присвоены таблице с помощью специальных символов (id_entry, iform); 2) аффиксы с id_dims связаны с ID_Word, что приводит к образованию слова в поле List of derivate

```
SELECT sg_form.name
FROM sg_entry, sg_form, coord_pairs
WHERE sg_entry.name='давлат'
AND sg_form.id_entry=sg_entry.id
AND coord_pairs.id=sg_form.id_dims
AND coord_pairs.str_pairs LIKE '%эгалик+келишик%' – эгалик ва келишикдаги сўзшакл.
```

List of derivate: *давлатимни / давлатингга / давлатидан / давлатнинг / давлатларида / давлатимизни / давлатингизда / давлатнинг ...*

В поле LT wordform_set_coord программы введен режим аффикса. В частности, wordform_set_coord (лемма + *s_ya + sh_ya + sin_ya). Здесь лемма – основа, аффикс s_ya – в форме слова, аффикс sh_ya – для формирования

формы, sin_ya – аффикс для формирования формы синтаксиса. Знак « \times » представляет ситуацию, когда метод вложения не является строго обязательным. Тогда форма слова появляется в тексте: `wordform_refresh` (`wrd`). Это именно то, что МА должны делать в машинном переводе.

Основные единицы в МА: единицы плана выражения (графемы); небольшие единицы значения (морфология); значимые части (слова); были проанализированы коммуникативные носители (предложения).

Тексты, написанные на узбекском языке, которые должны быть приняты во внимание на первом этапе для достижения ААР в различных секторах, рассматриваются в контексте МА, и были сделаны конкретные выводы относительно лингвистического обеспечения редакционной программы.

Алгоритм текстового АТТ. Как известно, любая программа работает на алгоритмах, а алгоритмы основаны на источниках данных. МА алгоритм использует следующие символы:

Us – база данных словарного запаса ЛО, $Us="SELECT * FROM 'Us'"$;

Ys – все встроенные аффиксы на узбекском языке, $Ys="SELECT * FROM 'Ys'"$;

Sq – база грамматических категорий на узбекском языке, $Sq="SELECT * FROM 'Sq'"$;

S_i – W словоформы извлеченные из текста, $1 \leq i \leq n$, n – W количество слов в тексте;

Sq_j – Sq аффиксы базы, $1 \leq j \leq m$, m – Sq количество аффиксов в базе

Tz – специальная функция, которая визуализирует неправильное написание слова и обеспечивает оптимальные параметры речи для слова с ошибкой.

Как только источник или база данных найдены, они будут взяты из аффилированного идентификатора вероятностного идентификатора, связанного с ним, на узбекском языке. Таким образом, алгоритм анализа заключается в следующем:

1. Каждое слово в S_i должно быть найдено в базе данных Us . Если найдено, переходите к следующему шагу, в противном случае шаг 5;

2. Получить (серийный номер) слова в базе Us ;

3. Искать аффикс Sq_j – Sq , который соответствует к номерам ID;

4. Если $S_i + Sq_j$ правильный, перейдите к шагу 10, в противном случае перейдите к следующему шагу;

5. Найдите каждое слово в S_i базе данных Ys . Если он найден, переходите к следующему шагу, в противном случае к шагу 3;

6. Получить базовый ID идентификатор S_i в базе Ys (серийный номер)

7. Посмотреть на аффикс базы данных Sq_j – Sq , соответствующий слову ID;

8. Если $S_i + Sq_j$ правильный, перейдите к шагу 10, в противном случае перейдите к следующему шагу;

9. Включите и выполните TZ и перейдите к шагу 10;

10. Проверьте правильность слова и перейдите к следующему слову $S+1$.

Мы анализируем рабочий порядок алгоритма на примере: *истеъмолчиларнинг* форма является основополагающей, то есть *wordform_set_coord* разбивается в стиле аффиксов, указанных в поле: *истеъмол/чи/лар/нинг*. После этого лемма потребления проверяется из базы данных List of Words, а аффиксы *-чи*, *-лар*, *-нинг* находится в базе данных Sqj–Sq, чтобы найти аффикс, нужно будет обратиться к базе данных Ys. Когда леммы и аффиксы находятся в ЛО то подтверждается, что слово написано правильно, а затем переходит к следующему слову.

Вторая часть «Технология анализа омонимов» относится к широко используемым методам определения омонимов в тексте. Был проанализирован метод Бриля, модель секретной Марковской модели, модификация моделей, создан лингвистическая технология определения омонимии в узбекских текстах. На основе разработанных источников и формул создана формула проверки лексем в текстах узбекского литературного языка.

S – Сборник текстов в официальном и научном стиле.

T_c – S количество меток в корпусе текста.

w_i – слово i – место в тексте, t_j – w в корпусе S . n – t_j количество тегов.

Слово – w_i в тексте и полученный $w_i + 1$ принимаются за биграмы.

$W_i = \{w_i, w_{i+1}\}$ – биграма, состоящая из i –слова в тексте и последующих слов.

$Y_j = \{w(t_j), w(t_{j+1})\}$ – S_j – в форме j – на корпусе j – биграма состоящий из следующих слов.

Теперь проверка слова – w_i производится по следующей формуле, основанной на наборе текстов – S , к которым правильно прикреплены аффиксы.

$$\prod_{j=1}^n (W_i, Y_j) \Rightarrow W_i$$

Принимая во внимание неспособность узбекского национального корпуса и отсутствие комплекса официальных и научных текстов, анализ омонимов на узбекском языке был основан на анализе N-граммы. Для этого в узбекский язык был введен лингвистическое снабжение по ID, что дает возможность анализировать грамматические образования контекста. Чтобы повысить точность анализа, виды омонимформ были идентифицированы и получены символические значение. После этого депозитная база была привязана к Models of CompWords (List of homonym=> Models of CompWords). В результате получается Q-аффикс, соответствующий номеру идентификатора M–Deposit на ЛО, в зависимости от N – слова, поставляемого с M –Deposit. То есть $M [id] + Q [id]$.

Мы анализируем этот метод на примере:

$N_{\text{com-ни}} + V$ (*отрезать рукава*), $N_{\text{com-га}} + V$ (*сунуть в рукава*), $\text{Pron}_{\text{ни(дир)}} + V$ (*всех победить*), $\text{Pron}_{\text{дан(дир)}} + V$ (*проиграть от меня*), $\text{Adj}_{\text{дан}} + V$ (*Vh*) (*проиграть от слабога*) и т.д.

Как правило, метод анализа омонимов объясняется тем, что определенное слово связано со словом через определенный грамматический индекс, что увеличивает точность анализа.

В разделе «**Модули программ автоматического анализа синтаксиса**» этой главы даны синтаксические стандарты. Как известно, объем предложения бесконечен и зависит от индивидуальной композиции или использования. По этой причине синтаксический анализ основан на вычислительных выражениях, т.е. в процессе создания модулей МА все лексемы были смоделированы в слова. В результате было создано 14 моделей словарного запаса, 16 моделей связности, 35 моделей ассоциаций управления в официальных и научных текстах.

Принимая во внимание тот факт, что каждый символ учитывается, и комбинацию меток на этикетке, предложение в лингвистическом обеспечении выглядит следующим образом:

Правило: в синтаксическом анализаторе слово является частью текста между двумя индикаторами (знаками препинания). Эти знаки следующие: точка / восклицательный знак / вопросительный знак + пробел и точка / восклицательный знак / вопросительный знак.

Схема:

[. / ! / ?] + [пробел] [. / ! / ?]

Пример:

[. / ! / ?] + [пробел] *Сувни тежаб ишлатиш лозим.*

Описан алгоритм автоматического синтеза анализа предложения.

В диссертационной работе алгоритм фонологических, морфологических и орфографических правил был создан для формирования лексико-грамматического кода в словарных модулях узбекского языка для редактирования и анализа узбекских текстов в разделе «Модули ключевых слов». В список слов узбекского литературного языка и в раздел «список слов» программы ААР были включены следующие: более 3000 существительных более 4000 глаголов, 1254 прилагательных, 488 наречие, 175 местоимение, 142 слога, 95 восклицательных, 63 модальные, 94 вспомогательные, 47 привязок, 36 частиц и 114 числительных.

В разделе, озаглавленном «Модуль формирования существительных», моделирование существительных дало языковые и синтаксические формы аффиксов, прикрепленных к их комбинациям. Известно, что слова с характеристиками изменения и владения объединяются под именительном падеже. Они включают существительных, прилагательных, числа, местоимения, слоговые слова, глаголы и названия движений. В исследовании лингвистические явления, происходящие, когда существительные объединяют синтаксические формообразные аффиксы, были расположены в соответствии с требованиями КЛ в следующих темах:

“Орфографик қоидалар”, “-к / -қ товушлари билан тугайдиган сўзларга эгалик аффиксларининг қўшилиши алгоритми”, “Эгалик аффикслари қўшилганда асосда товуш туширилиши алгоритми”, “Парадигматик ва нопарадигматик шаклларининг кетма-кет келиши меъёрлари”, “Феъл

шаклининг умумий ясашиш қоидаси”, “*Феъл шаклини ҳосил қилишда морфларнинг мос келиши ва мос келмаслик қоидаси*”, “*Морфонологик ва фонологик қоидалар*” ва “*Истисно қоидаси*”.

Последний имеет свою уникальную форму, а количество видов регулируется в разделе «**Модуль формирования чисел**».

В «**Модуле аффиксации в лингвистическом процессоре**» диссертации приводится метод прикрепления из четырех части речи (существительное, прилагательное, глагол, наречие).

Лексема или лемма, которая присоединяет аффикс в базе данных, называется реализацией лексемы.

В узбекском литературном языке к лексису добавляются словообразные суффиксы в следующем порядке: (префикс₁) + основа₂ + (составленный аффикс₃) (*хуш₁хабар₂; қабул₂ хона₃; бе₁одоб₂лик₃*)

Лингвистическое обеспечение программы включает в себя все существующие языковые аффиксы на узбекском языке. Это поможет улучшить будущее аналитическое программное обеспечение и автоматически анализировать новые слова.

На узбекском литературном языке выявлено 114 аффиксов. Из них 37 – существительных, 11 – прилагательные, 54 – глаголы, 3 – наречий, 1 – числительное, 1 – местоимение, 1 – восклицательная и 1 – модальная лемма.

Всего было обнаружено 58 аффиксов. Из них 17 существительных 16 прилагательных, 3 числительных, 7 наречий, 3 местоимений, 3 модальных лемм.

На узбекском языке было выявлено 117 появлений аффиксов. Из них 56 существительных, 40 глаголов, 11 прилагательные, 3 наречие, 2 местоимение, 2 модальных и 3 примера леммы создают новую прилагательную лемму.

На основе источников было обнаружено 48 аффиксов. Из них 18 существительных 16 прилагательных, 5 глаголов, 4 наречие, 3 местоимение и 2 модальные которые присоединяются и производят леммы наречий.

ЗАКЛЮЧЕНИЕ

При создании программы лингвистического обеспечения редактирования и анализа узбекских текстов были сделаны следующие выводы:

1. Исследовательские работы, которые были проведены до сих пор, были частично сфокусированы на автоматическом редактировании и анализе. В современных исследованиях два языка, морфология языков были проанализированы сравнительно типологически. Тем не менее, этот вопрос можно рассматривать как лингвистическую систему, которая давно интересовала лингвистов.

2. Всемирное исследование компьютерной лингвистики, в свою очередь, приводит к текстовому редактированию и анализу, помогая им делать выводы и результаты, которые актуальны и надежны. Реализация этих идей приведет к новому уровню изучения текстов в компьютерной лингвистике и облегчит ее практическое применение.

3. Создание лингвистической поддержки программного обеспечения для автоматического анализа и редактирования было определено разработкой лингвистических правил и правил с алгоритмом, введением лексикографических ресурсов, структурой моделей взаимосвязи слов.

4. В этом процессе официальные и научные тексты не редактируются и анализируются лингвистически, а разрабатываются принципы написания текстов и основанные на грамматике системы или программное обеспечение. Трудно сказать, что слово может быть использовано в стиле и учитывая стандарты этих методов, сложно создать форму выражения со всем его стилем и внешним видом. Поэтому использование научных и формальных методов речи и аффиксов при создании лингвистической поддержки программного обеспечения для автоматического анализа и редактирования поможет повысить эффективность работы.

5. Источником лингвистической базы любой программы является «Морфологический словарь узбекского языка». Морфологический словарь экономит время лингвиста и ускоряет достижение результата. Этот словарь содержит некоторые термины, на которых написана лексема, и его произношение. Это обеспечивает идеальный результат программы. Поэтому желательно разработать словарь, чтобы обеспечить удобство создания лингвистического обеспечения программ обработки текста на узбекском языке и быстрого достижения результатов.

6. Роль и значение лингвистического модуля в создании лингвистического процессора имеет решающее значение. В компьютерной лингвистике термин модуль применяется к независимому компоненту программного обеспечения, охватывающему определенное лингвистическое программирование. В лексическом модуле, если слой словаря окружен, в графематическом модуле редактируются символы, знаки препинания, алфавит и другие символы, правила орфографии в орфографическом модуле, синтез слов в морфологическом модуле (анализ слова в слово) и синтез

(процесс грамматического анализа лексемы), в синтаксическом модуле суперсинтаксическая единица анализируется взаимодействием слов.

7. Желательно модулировать графические, морфологические и синтаксические этапы для эффективного функционирования и эффективности лингвосистемы, чтобы редактировать и анализировать официальные и научные тексты при первоначальном введении ААР узбекского языка. Определение единиц, которые будут выполнены на графическом этапе, обеспечит точность редактирования. Принципы морфологического анализа на примере существительных позволяют создать принципы создания других модулей. В синтаксическом анализе количество речи бесконечно, и лучше полагаться на слова, которые являются относительно небольшими единицами в СА, потому что это зависит от индивидуальной структуры или использования. Таким образом, в процессе создания модулей МА все лексемы разделяются на слова, моделируется слово и создается «категория словосочетаний».

8. Орфографические, фонологические, морфологические явления встречаются в словообразовании. При создании лингвистического модуля лингвистические стандарты современного узбекского литературного языка рассматриваются как основное языковое явление в анализе и анализе текста. Создание алгоритма фонологических, морфологических и орфографических правил с целью формирования лексиконного грамматического кода в словарных модулях узбекского языка в компьютерной лингвистике на основе традиционных лингвистических норм; создание лингвистической модели синтаксического и семантического кода, предоставление грамматических правил и создание их алгоритма. Этот вопрос не только обеспечивает качество лингвоанализа, но и показывает актуальность узбекского языка.

9. Статистический расчет аффиксов на узбекском языке и создание их лексической базы и модели лемма + аффикс являются важной составляющей языкового предложения. Для компьютера важно анализировать аффиксы отдельно в форме аффиксообразующих и формирующих слов, идентифицируя их модели.

10. Омним формы всегда считались одним из событий в языковом центре. Поэтому тексты изучаются отдельно в русской и зарубежной лингвистике. Метод идентификации или исключения текстовых слов, которые были разработаны до сих пор, был изучен методом Бриля, секретной Марковской моделью, N-граммой и другими способами, а также методом анализа живых форм в узбекских текстах. В результате была создана основа для анализа омним форм.

11. Создание категориальных моделей представляет собой сложную проблему, поскольку фразеологические соединения не являются речевыми, а лингвистическими, их грамматические связи устойчивы. По этой причине не только модель ссылок, но и грамматическая форма, а также включение лингвистической базы данных обеспечивает точный автоматический анализ.

12. Программное обеспечение для лингвоанализа важно не только как инструмент для ускорения процесса редактирования и анализа, но и как

инновационная система, которая сочетает в себе способность писать грамотные тексты на узбекском языке.

13. Всестороннее изучение лингвистических схем, которые были собраны до настоящего времени, внесет важный вклад в создание узбекского лингвистического корпуса в будущем.

**SCIENTIFIC COUNCIL AWARDING SCIENTIFIC DEGREES
DSc.27.06.2017.Fil.05.02. AT THE FERGANA STATE UNIVERSITY**

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

ABZHALOVA MANZURA ABDURASHETOVNA

**LINGUISTIC MODULES FOR THE PROGRAM OF EDITING AND
ANALYZING TEXTS IN THE UZBEK LANGUAGE
(FOR THE PROGRAM OF ANALYZING OFFICIAL AND SCIENTIFIC
TEXTS)**

10.00.11 - Theory of language. Applied and computational linguistics

**DISSERTATION ABSTRACT OF THE DOCTOR OF PHILOSOPHY
(PHD) IN PHILOLOGICAL SCIENCES**

Fergana – 2019

The theme of the doctoral thesis is registered by Supreme Attestation Commission under the Cabinet of Ministers of the Republic of Uzbekistan as № B2019.2.PhD/Fil.320

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

The Abstract of thesis is written in 3 languages (Uzbek, Russian, English (Summary)) and is placed on the website (www.fdu.uz) of the Scientific Council and Information-educational portal “Ziyonet” (www.ziyonet.uz)

Scientific supervisor: **Dadaboev Hamidulla Aripovich**
Doctor of Philology, Professor

Official opponents: **Ulukov Nosirjon Muhammadalievich**
Doctor of Philology, Professor

Pardaev Azamat Bakhronovich
Doctor of Philology

Leading organization: **Bukhara State University**

Defense of the dissertation will take place on “_____” _____ 2019 at “_____” at the meeting of the Scientific Council awarding Scientific degrees DSc.27.06.2017.Fil.05.02. at the Fergana State University (Address: 100151 Fergana city, Murabbiylar Street, 19. Phone: 244-44-29. Fax. (+99873) 244-66-03, e-mail: fardu.info@umail.uz.

Doctoral dissertation could be viewed at the Information Resource Center of the Fergana State University (registration # _____). Address: 100151 Fergana City, Murabbiylar Street, 19. Phone (99873) 244-44-02.

Abstract of dissertation is mailed out on “_____” _____ 2019
(Protocol at the Register # _____ from “_____” _____ 2019)

M.Hakimov

Chairman of the Scientific Council awarding Scientific degrees, Doctor of Philology, Professor

M.T.Zokirov

Scientific secretary of the Scientific Council awarding scientific degrees, PhD in Philological sciences, assistant professor

S.Muminov.

Chairman of the Scientific seminar at the Scientific Council awarding scientific degrees, Doctor of Philology Sciences, Professor

INTRODUCTION (abstract of PhD thesis)

The aim of the research is to create the linguistic content of official and scientific texts of the Uzbek literary language at the lexical, morphological and syntactic levels of computer literary and linguistic modules of the analytical program and the computer information corpus that unites the linguistic complex of the Uzbek language.

As an object of study, texts written in official and scientific texts in the Uzbek literary language were chosen.

The subject of the research is the development of the basics of language support for automated editing of Uzbek texts and analysis software.

Scientific novelty of the research:

for the first time in the Uzbek CL a theoretical and scientific basis was created for the linguistic support of the AAE system and the linguistic, practical and educational value of the AAE system was disclosed;

the principles of linguistic support AAE, as well as the module and modulation in the field of computational linguistics;

technology of texts and AAE modules for graphematic editing, stages of morphological and syntactic analysis, which ensure the superiority of the AAE program;

methods for determining the stability of different words in different families and the method for analyzing random forms in Uzbek texts;

linguistic support of the program, principles of the program linguistic base and the algorithm of morphological analysis.

Implementation of research results. The results of the drafting of official and scientific text editorials obtained in the dissertation research were presented in the following areas:

The thesis was used in “Creating multilingual dictionaries of electronic and banking and financial terminology related to the Uzbek language and its translation” as a result of translation of vocabulary and statistics and a formal approach to the Uzbek language. (Reference No. 89-03-915 of the Ministry of Higher and Secondary Special Education of March 5, 2019) Multilingual dictionaries were improved based on scientific results;

The analysis and conclusions on the syntax analysis using AAE programs and phrases models were used in the Annex to the project of state grants on the topic “Creating audiobooks on linguistics subjects for students with disabilities”. (Reference No. 89-03-915 of the Ministry of Higher and Secondary Special Education of March 5, 2019) As a result, a system of laboratory assignments was developed for the syntax section of the Modern Uzbek literary language.

The dissertation materials were used on the TV channel “Navoi” for the program “Daxldorlik Tuyg’usi” and the radio channel “Uzbekistan” for the program “Til - Millat Ko’zgusi”, “Talim va islohot”. (Information No. 01-02 / 57 of the Navoi Regional Television and Radio Company of March 11, 2019). As a result, the scientific potential of broadcasting and broadcasting was ensured.

Structure and size of the dissertation. The dissertation consists of an introduction, three chapters, a conclusion and a list of used literature. The total amount of work is 164 pages.

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

1. Abjalova M.A. The linguistic norms of the grapheme editing module //International Journal of Science and Research (IJSR), India. ISSN: 2319-7064. Volume 8 Issue 2, 2019. – P. 1005-1007. (№ 40; Research Gate: 0.28) www.ijsr.net.

2. Абжалова М.А. Ўзбек тилидаги матнларни таҳрир қилишда услуб масаласи // Таълим тизимида ижтимоий-гуманитар фанлар, – Тошкент, 2013. – №2. – Б. 140-145. ISSN 2181-7286. 1.01.2011 (10.00.00 №16)

3. Абжалова М.А. Матнларни автоматик таҳрир ва таҳлил қилиш дастурининг лингвистик таъминоти манбалари // Илм сарчашмалари. – Урганч, 2015. – №6. – Б. 36-39. ISSN 2010-6246. (10.00.00. № 3)

4. Абжалова М.А. Матнни графематик таҳрирлаш модулининг лингвистик нормалари// Илм сарчашмалари. – Урганч, 2018. – №3. – Б. 151-156. ISSN 2010-6246. (10.00.00. № 3)

5. Особенности стилей узбекского языка в программе автоматического анализа текста // Научное сообщество студентов XXI столетия. Гуманитарные науки:материалы XI студенческой международной заочной научно-практической конференции. (14 мая 2013 г.) – Новосибирск: СибАК, 2013. – С. 310-316.

6. Абжалова М.А. Об автоматическом морфологическом анализе узбекского языка / Глобал олий таълим тизимида илмий тадқиқотларнинг замонавий услублари. Халқаро илмий-амалий анжуман тўплами. – Навоий, 2015. – Б. 90-92.

7. Абжалова М.А. Матнларни автоматик таҳрир ва таҳлил қилувчи дастур босқичлари / XXI аср – интеллектуал авлод асри. Ёш олим ва талабаларнинг Республика илмий-амалий анжумани материаллари. – Тошкент, 2015. – Б. 7-11.

8. Абжалова М.А. Матнларга автоишлов бериш тизимлари / VI th International conference on computer processing of Turkic languages TURKLANG – 2018. – Tashkent, 18-20 october, 2018. – С. 243-245.

II бўлим (II часть; II part)

9. Абжалова М.А. Программа автоматического синтеза и анализа текстов узбекского языка. № DGU 20190621 рақамли илмий гувоҳнома, 2019.

10. Абжалова М.А. Эсперанто тилини биласизми? – Тошкент: Mumtoz soʻz, 2010. – 64 б. ISBN 978-9943-363-73-1

11. Абжалова М.А. Айрим аффиксларнинг ноодатий қўлланишига доир / Тилшуносликка илк қадам. Илмий мақолаларнинг III тўплами. – Тошкент, 2007. –Б. 44-46.

12. Абжалова М.А. Ўзбек тилини формаллаштиришда эсперанто тилининг ўрни / Тилшуносликка илк қадам. Ўзбекистон Миллий университетининг 90 йиллигига бағишланган махсус тўплам. – Тошкент: Университет, 2008. – Б. 98-101.

13. Абжалова М.А. Матнни морфологик даражада автоматик таҳрир қилиш / Лингвист. Илмий мақолалар тўплами. 1-жилд – Тошкент, 2010. – Б. 87-90.

14. Абжалова М.А. Лингвистик таҳлил дастурлари ва матнни таҳрир қилиш / Лингвист. Илмий мақолалар тўплами. 1-жилд – Т. 2010. – Б. 97-101.

15. Абжалова М.А., Тошимов Р.Ю. Матнларни лингвистик таҳрирлашнинг психологик асослари / Ўзбекистоннинг мустақил тараққиёт ва ижтимоий-фалсафий тафаккур янгиланиши. Ёш олимлар илмий конференцияси материаллари. II қисм – Тошкент, 2011. – Б.155-158.

16. Абжалова М.А. Ўзбекча матнларни автоматик таҳрир қилиш дастурининг таъминотини яратиш тамойиллари / Магистрларнинг илмий мақолалар тўплами. – Тошкент: ЎзМУ. 2011.

17. Абжалова М.А. Тилнинг софлиги – миллий баркамоллик мезони / Лингвист. Илмий мақолалар тўплами. 2-жилд – Тошкент, 2011. – Б.129-132.

18. Абжалова М.А. Эсперанто халқаро тили ҳақида қисқача маълумот. Компьютер лингвистикаси дарслигининг 20-параграфи. – Тошкент: Akademnashr, 2011. – Б. 242-365.

19. Абжалова М.А. Компьютер лингвистикасида лингвистик модуль тушунчаси / Лингвист. Илмий мақолалар тўплами. 3-жилд. – Тошкент, 2012. – Б.133-137. ISBN 978-9943-389-49-6

20. Абжалова М.А. Автоматик таҳрир дастурини яратишдаги лингвистик муаммолар / Современные технологии и инновации горно-металлургической отрасли. Материалы Республиканской научно-технической конференции. – Навои, 2012. – Б. 478-479.

21. Абжалова М.А. Автоматический морфологический анализ узбекского языка / XXI аср – интеллектуал авлод асри. Худудий илмий-амалий анжуман тўплами. – Навоий, 2014. – Б.128-130.

22. Абжалова М.А. Автоматик таҳрир тизимининг НДКИ хузуридаги 2-сон академик лицейи ўқувчилари ўртасида тарғиб қилиниши / Академик лицейлардаги таълим-тарбия жараёнини янги босқичга кўтариш ва такомиллаштиришда ўқитувчилар салоҳиятини ошириш. Республика ўқув-амалий семинар тўплами. – Навоий, 2014. – Б. 49-50.

23. Абжалова М.А. Матнларни автоматик таҳрир қилишдаги айрим муаммолар ва уларнинг ечими / Лингвист. Илмий мақолалар тўплами. 6-жилд – Тошкент, 2015.

24. Абжалова М.А. Матнларни автоматик таҳрир ва таҳлил қилувчи дастурнинг лингвистик босқичлари / XXI аср – интеллектуал авлод асри. Худудий-илмий конференция тўплами. – Навоий, 2015.

25. Абжалова М.А. Расмий услубдаги матнларни автоматик таҳрирлаш имконияти / Ўзбек лексикографиясининг долзарб масалалари. Илмий мақолалар тўплами. – Тошкент: ТДЎТАУ, 2017.

26. Абжалова М.А. Матнларга автоматик қайта ишлов берувчи дастурларнинг лингвистик таъминоти манбалари / Ўзбек терминологияси: бугунги ҳолати ва истиқболи. Илмий мақолалар тўплами. – Тошкент: ТДЎТАУ, 2017.

27. Абжалова М.А. Матнларни автоматик таҳрир ва таҳлил қилиш дастурининг лингвистик модуллари / Ўзбек тилшунослигининг долзарб масалалари. – Тошкент: ТДЎТАУ, 2018. – Б. 104-106.

28. Абжалова М.А. Ягона корпоратив автоматлаштирилган ахборот кутубхона тизимини яратиш – хизмат кўрсатиш сифатини ошириш омилидир / Фан, таълим, маданият ва бизнесда ахборот-кутубхона ресурслари. “Central Asia – 2018” XII халқаро конференция тўплами. – Сурхондарё, 2018. – Б. 93-97.

29. Абжалова М.А. Саводхонликни оширишда матнларни автоматик таҳрир қилиш тизимининг ўрни / Тил ва адабиёт таълимида замонавий ахборот ва педагогик технологиялар. Республика илмий-амалий анжуман тўплами. – Тошкент: ТДЎТАУ, 2018. (электрон нашр)

30. Абжалова М.А. Автоматическая система анализа // Monografia pokonferencujna. Science, research, development #13. Technics and technology. – Berlin (Берлин). 30.01.2019 - 31.01.2019. – P. 125-127.

31. Абжалова М.А. Илмий тадқиқот ишини олиб боришда амалий ишланмаларнинг аҳамияти ва уларни кутубхоналарда сақлаш масаласи / Фан, таълим, маданият ва бизнесда ахборот-кутубхона ресурслари. “Central Asia – 2019, Urganch” XIII халқаро конференция тўплами. – Тошкент, 2019. – Б. 163-167.

Автореферат «FarDU. Ilmiy xabarlar – научный вестник ФерГУ» журналі тахририятида тахрирдан ўтказилди (23.04.2019 йил).

Босишга рухсат этилди: 10.06.2019 йил
Бичими 60x84 ¹/₁₆. «Times New Roman»
гарнитурда рақамли босма усулда чоп этилди.
Шартли босма табоғи 3,0. Адади 100. Буюртма № 26

Ўзбекистон Республикаси Фанлар Академияси
Асосий кутубхонаси босмахонасида чоп этилди.
Тошкент шаҳри, Зиёлилар кўчаси, 13-уй.

