

**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)диссертация
автореферати мундарижаси**

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

**Contents of dissertation abstract of Doctorof 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

Эълонқилинганишларрўйхати

Список опубликованных работ

Listofpublications.....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**

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.ziynet.uz) жойлаштирилган.

Илмий раҳбар:

Дадабоев Ҳамидулла Арипович
филология фанлари доктори, профессор

Расмий оппонентлар:

Улуков Носиржон Муҳаммадалиевич
филология фанлари доктори, профессор

Пардаев Азамат Бахронович
филология фанлари доктори

Ётақчи 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.Ҳақимов
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 тўғрисида”ги, 2018 йил 5 июндаги

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

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 меъриё-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишга ушбу диссертация иши муайян даражада хизмат қилади.

Тадқиқотнинг республика фан ва технологиялари ривожланиши-нинг 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.Чижиковский, Г.Г.Белоногов, И.С.Дуганова, Л.Л.Иомдин, Э.К.Лавошникова,

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.

³ 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.

⁴ 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://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.

А.В.Сокирко, И.М.Ножов, Е.И.Большакова, 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.Маҳмудов, 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 М.Х.Ҳакимов раҳбарлигида

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. – 160b.; 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 T., Po'latova I.O.

“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 илмийуслугларидаги 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 таҳлилинини 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 алгоритминини 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 таҳрирлаш ва

Dunyoviy o‘zbek tili (O‘zbek tilida fe’lning 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.

¹¹Хақимов М.Х. Расширяемый входной язык математического моделирования естественного языка для многоязычной 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.

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. Мавзунини 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 таъминотинини 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 мавзуйий 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, таҳлил бирликлари, табиий тилдаги

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.

¹⁶ Апресян, 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.

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 тартибда 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 перефраз (тасвирий ифода), 1393 фразема базаси яратилди. Шунингдек, матнда мавжуд омонимшакллар таҳлилини 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 ривож 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.

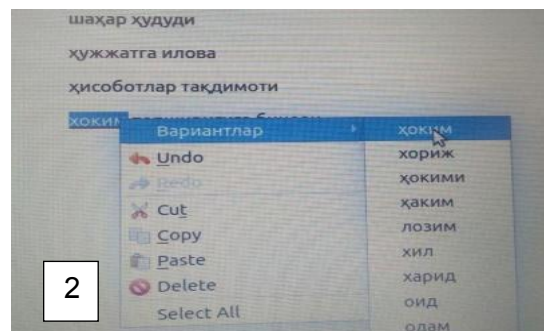
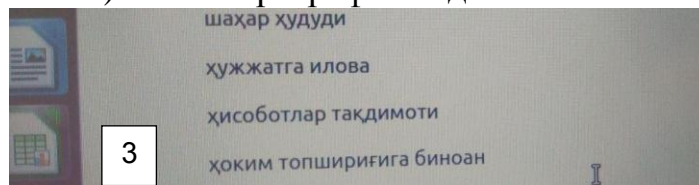
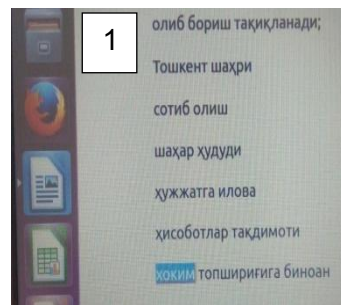
²¹Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учебное пособие. – Москва: МИЭМ, 2011. – С. 112-122.

Фикримизга дастурдаги фрагментлар билан аниқлик киритамиз:

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

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

3) хато бартараф 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’, ‘муסיқа’, ‘фестивалига’, ‘тайёргарлик’, ‘кўриш’, ‘ва’, ‘уни’, ‘ўтказиш’, ‘тўғрисида’, ‘»’, ‘зи’, ‘ПҚ’, ‘-’, ‘103’, ‘-’, ‘сонли’, ‘қарори’, ‘ўз’, ‘кучини’, ‘йўқотган’, ‘деб’, ‘ҳисоблансин’, ‘.’]

Бу ерда квадрат қавс матннинг бошланиш ва яқунланиш жойини кўрсатади. Биртирноқ элементни ажратувчи белги, вергул ажратилган белгиларни бир-

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

биридан алоҳидалашга хизмат қилади. Мазкур мисолда токенизация жараёнида 'зи' бўғини маънога эга бўлмагани боис кесиб 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); ҲРБ – 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 қуйидаги ишлар амалга оширилиши талаб қилинади: 1) тилнинг луғат таркибини компьютерга киритиш, яъни электрон луғатни тузиш; 2) луғатдаги адабий тилга мансуб сўзларни ажратиб олиш; 3) ажратиб олинган сўзларни услубий (илмий, бадий, 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)

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

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

тилшунослар томонидан амалга оширилган юқоридаги ишлар дастурчи 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 кўпликка эга сўзлар (асосан, саналмайдиган отлар, мавҳум отлар) sol_CorrNounNumber турухига 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_Word га боғланади, натижада List of derivate майдонида сўзшаклини 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(wrd). 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, $U_s = \text{"SELECT * FROM 'Us'"};$

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

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

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 $ID_{си}$ 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 кейинга 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.

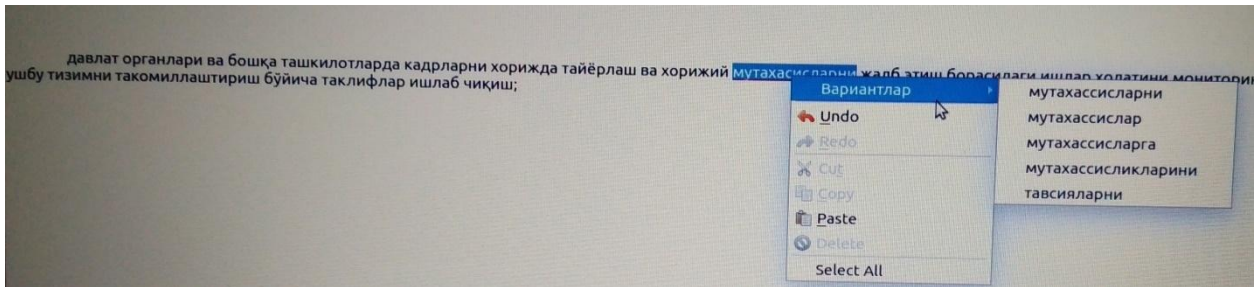
10. Сўз тўғри ёзилганлиги тасдиқлансин ва кейинги S_{i+1} сўзга ўтилсин.

Алгоритм иш тартибини мисол орқали таҳлил қиламиз: *истеъмолчиларнинг* сўзшакли стемминг қилинади, яъни `wordform_set_coord` майдонида берилган аффикслар тартиби бўйича кесиб чиқилади: *истеъмол/чи/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 сўзларшаклларни 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 куйидаги 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 омонимшаклларни таҳлил қилиш технологияси N-grammга боғлиқ равишда яратилди. Бунинг учун ўзбек тилидаги омонимлар ЛТга ID билан киритилди, бу

жихат 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: $N_{\text{com-ни}} + V(\text{енгини кесмоқ}), N_{\text{com-га}} + V(\text{енгига қадамоқ}), \text{Pron}_{\text{ни(дир)}} + V(\text{хаммани энгмоқ}), \text{Pron}_{\text{дан(дир)}} + V(\text{мендан энгилмоқ}), \text{Adj}_{\text{дан}} + V(\text{Vh})(\text{кучсиздан энгилмоқ})$ 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 боғланишининг 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: нуқта / ундов белгиси / сўроқ белгиси + бўшлиқ ва нуқта / 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 куйидаги тартибда қўшилади: (префикс₁)+асос₂+(ясовчи аффикс₃)(хуш₁хабар₂; қабул₂хона₃; бе₁одоб₂лик₃)

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

Ўзбек адабий тилида аффиксация усули билан от ясовчи 114та 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 куйидаги 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 ҳаминша 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 бекусур ёзиш кўникмасини шакллантирувчи инновацион система сифатида муҳим аҳамиятга эга ҳисобланади.

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