

ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМий ДАРАЖАЛАР БЕРУВЧИ
DSc.13/30.12.2019.T.07.01 РАҚАМЛИ ИЛМий КЕНГАШ

ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ

АБИДОВА ШАХНОЗА БАХОДИРОВНА

**“ТУРКИЙ ТИЛЛАР УЧУН ЭЛЕКТРОН ТАРЖИМАНИНГ
МОДЕЛЛАРИ ВА АЛГОРИТМЛАРИ (ЎЗБЕК ТИЛИДАН
ҚОРАҚАЛПОҚ ТИЛИГА ТАРЖИМА МИСОЛИДА)”**

05.01.04 – Ҳисоблаш машиналари, мажмуалари ва компьютер
тармоқларининг математик ва дастурий таъминоти

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

Тошкент – 2021

**Техника фанлари бўйича фалсафа доктори (PhD) диссертацияси
автореферати мундарижаси**

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

**Contents of dissertation abstract of doctor of philosophy (PhD)
on technical sciences**

Абидова Шахноза Баходировна

“Туркий тиллар учун электрон таржиманинг моделлари ва алгоритмлари
(Ўзбек тилидан қорақалпоқ тилига таржима мисолида)” 3

Абидова Шахноза Баходировна

«Модели и алгоритмы электронного перевода для тюркских языков (на
примере перевода с узбекского языка на каракалпакский язык)» 21

Abidova Shakhnoza Baxodirovna

"Models and algorithms for electronic translation for Turkic languages (on the
example of translation from Uzbek into Karakalpak)" 39

Эълон қилинган ишлар рўйхати

Список опубликованных работ

List of published works 43

АБИДОВА ШАХНОЗА БАХОДИРОВНА

**“ТУРКИЙ ТИЛЛАР УЧУН ЭЛЕКТРОН ТАРЖИМАНИНГ
МОДЕЛЛАРИ ВА АЛГОРИТМЛАРИ (ЎЗБЕК ТИЛИДАН
ҚОРАҚАЛПОҚ ТИЛИГА ТАРЖИМА МИСОЛИДА)”**

05.01.04 – Ҳисоблаш машиналари, мажмуалари ва компьютер
тармоқларининг математик ва дастурий таъминоти

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

Тошкент – 2021

Техника фанлари бўйича фалсафа доктори (PhD) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида В2021.1.PhD/Т2080 рақам билан рўйхатга олинган.

Диссертация Муҳаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университетида бажарилган.

Диссертация автореферати уч тилда (ўзбек, рус, инглиз (резюме)) Илмий кенгаш веб саҳифасида (www.tuit.uz) ва «Ziyonet» Ахборот таълим порталида (www.ziyonet.uz) жойлаштирилган.

Илмий раҳбар: **Нуралиев Фахриддин Муродиллаевич**
техника фанлари доктори, доцент

Расмий оппонентлар: **Мухамедиева Дилноз Тулкуновна**
техника фанлари доктори, профессор

Норов Абдисант Мурадович
техника фанлари бўйича фалсафа доктори (PhD)

Етакчи ташкилот: Ислом Каримов номидаги Тошкент Давлат техника университети

Диссертация ҳимояси Муҳаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университети ҳузуридаги DSc.13/30.12.2019.Т.07.01 рақамли Илмий кенгашнинг 2021 йил «29» 06 соат 9:00 даги мажлисида бўлиб ўтади. (Манзил: 100202, Тошкент шаҳри, Амир Темура кўчаси, 108-уй.

Тел.: (99871) 238-64-43, факс: (99871) 238-65-52, e-mail: tuit@tuit.uz).

Диссертация билан Муҳаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университети Ахборот-ресурс марказида танишиш мумкин (209 рақам билан рўйхатга олинган). (Манзил: 100202, Тошкент шаҳри, Амир Темура кўчаси, 108-уй. Тел.: (99871) 238-65-44).

Диссертация автореферати 2021 йил «15» июни куни тарқатилди.
(2021 йил «10» июни даги 17 рақамли реестр баённомаси.)



Р.Х. Хамдамов
Илмий даражалар берувчи илмий кенгаш раиси, т.ф.д., профессор

Э.Ш. Назирова
Илмий даражалар берувчи илмий кенгаш илмий котиби, т.ф.д., доцент

М.А. Рахматуллаев
Илмий даражалар берувчи илмий кенгаш қошидаги илмий семинар раиси, т.ф.д., профессор

КИРИШ (фалсафа доктори (PhD) диссертацияси аннотацияси)

Диссертация мавзусининг долзарблиги ва зарурати. Жаҳон тилшунослигида XX асрдан бошлаб компьютер лингвистикаси соҳасига оид терминлар тизимини яратиш, тартибга солиш, уларни илмий жиҳатдан тадқиқ этиш масаласини ўрганишга эътибор кучайди. Жамиятнинг ижтимоий-сиёсий, иқтисодий ва маънавий ҳаётида рўй берган улкан ўзгаришлар туфайли туркий тиллар оиласига мансуб терминларни машинали ўқитиш ва компьютер лингвистикаси нуқтаи назаридан ўрганиш, уларнинг компьютерли моделлаштириш жараёнларини амалга ошириш бугунги кундаги долзарб масалалардан бири ҳисобланади. Зеро, табиий тилнинг компьютерга йўналтирилган моделларини яратиш ва бу моделлар асосида бир қатор масалаларни ҳал этиш, табиий тилларни морфологик, синтактик ва семантик жиҳатдан автоматик таҳлил қилиш, ҳамда уларнинг ўзига хос жиҳатларини ёритиш компьютер лингвистикасида муҳим аҳамият касб этади. Компьютер лингвистикаси бўйича ривожланган мамлакатлар, жумладан Россия Федерацияси, Япония, Германия, АҚШ, Хитой, Англия, Корея, Франция, Австралия ва бошқа мамлакатларда бир тилдан иккинчи бир тилга таржима жараёнини амалга оширишда фойдаланилган усуллар, моделлар ва дастурий технологияларни ишлаб чиқишга алоҳида эътибор қаратилмоқда.

Жаҳонда классик назариялар базаси, ҳамда математик ва функционал таҳлилнинг зарур бўлимларини расмийлаштириш, ҳисоблаш математикасига асосланган замонавий дастурлаш технологиялари асосида теоремаларни исботлаш учун алгоритмик усулларни ишлаб чиқиш алгоритмлаштиришнинг долзарб муаммоларидан бири ҳисобланади. Савол-жавоб тизимларининг алгоритмик асослари табиий тилда маълумотларни манипуляция қилиш билан қурилган. Бу шунингдек, бир тилдан иккинчисига ўтиш билан боғлиқ алгоритмик дастурлаш муаммоси ва интеллектуал тизимларни яратиш учун алгоритмик асосларни қуриш ва бошқаларни ўз ичига олади. Алгоритмик банкларда ахборот қисмларини тўлдириш ва уларнинг операцион қисмлари учун дастурий таъминотни ишлаб чиқиш муҳим аҳамият касб этмоқда. Туркий тиллар оиласига мансуб ўзбек тилининг компьютерли лингвистикасини яратиш ва электрон таржимонлар фаолиятини оптималлаштириш долзарб вазифадир. Бу бизга ўзбек тилини компьютерга йўналтирилган моделларини яратишга имкон беради, яъни ўзбек тилини ўрганиш, билимларни баҳолаш, ўзбек тилидаги матнларни синаб кўриш, компьютерга матнларни овозли киритиш, матнларни таҳрирлаш масаласига алоҳида эътибор қаратиш зарурлигини англатади.

Республикамизда икки тилли электрон таржимонларни яратиш, соҳага оид сўзлар учун маълумотлар базасини қуриш уларнинг изоҳли матнлари корпусини яратиш бўйича илмий тадқиқотлар олиб борилмоқда. Бунга мисол тариқасида ўзбек тилининг лингво-мантикий модели уларнинг компьютер лингвистикасини шакллантириш, ўзбек тилининг соҳаларга оид миллий корпуси ва маълумотлар базасини шакллантириш уларнинг электрон луғатларини ишлаб чиқиш, туркий тиллар ва бошқа тиллар ўртасида

автоматик таржима моделларини ишлаб чиқиш ва бошқаларни келтириш мумкин. 2017-2021 йилларда Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегиясида «...миллий контентни ривожлантириш, давлат тилидаги таълим, илмий-маърифий, ёшлар эҳтиёжларига мос замонавий ахборот ресурсларини, мультимедиа маҳсулотларини яратиш ва тарғиб қилиш механизмларини такомиллаштириш, ахборот-коммуникация технологияларини жорий этиш» вазифалари белгиланган. Бу вазифаларни амалга оширишда ўзбек лингвистикасининг компьютерга йўналтирилган моделларини яратиш, ҳозиргача ўзбек компьютер лингвистикасида машина таржимасига оид тадқиқотларнинг етарли эмаслиги сабабли ўзбек тилидан хорижий тилга ва аксинча таржима дастури учун ўзбек тилининг лингвистик маълумотлар базасини яратиш зарур ва булардан амалда фойдаланиш учун алгоритмлар ва дастурий воситаларни яратиш муҳим масалалардан бири ҳисобланади.

Ўзбекистон Республикаси Президентининг 2019 йил 21 октябрдаги ПФ-5850-сон «Ўзбек тилининг давлат тили сифатидаги нуфузи ва мавқеини тубдан ошириш чора-тадбирлари тўғрисида»ги Фармонида «...давлат тилининг ахборот ва коммуникация технологиялари, хусусан, Интернет жаҳон ахборот тармоғида муносиб ўрин эгаллашини таъминлаш, ўзбек тилининг компьютер дастурларини яратиш» вазифалари белгиланган ҳамда мазкур фаолиятларга тегишли меъёрий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишда ушбу диссертация тадқиқоти муайян даражада хизмат қилади.

Тадқиқотнинг республика фан ва технологиялар ривожланишининг устувор йўналишларига мослиги. Мазкур тадқиқот республика фан ва технологиялар ривожланишининг IV. «Ахборотлаштириш ва ахборот-коммуникация технологияларини ривожлантириш» устувор йўналишлари доирасида бажарилган.

Муаммонинг ўрганилганлик даражаси. Жаҳонда электрон таржимонларнинг модел ва алгоритмларини ишлаб чиқиш бўйича жуда кўп тадқиқотлар олиб борилган бўлиб, жумладан, электрон таржимон асосчиси У.Уивер, Р.Риченс, А.А.Ляпунов, Г.Г.Белоногов, В.Г.Васильев, Л.А.Гращенко, Ю.Н.Марчук ва бошқаларнинг бу борадаги ишлари қиёсий ўрганиб чиқилди. Ушбу йўналиш бўйича параллел тадқиқотлар А.И.Герцен номидаги Ленинград Давлат педагогика институтининг муҳандислик Лингвистика лабораториясида олиб борилган илмий тадқиқот ишлари қиёсий ўрганиб чиқилди.

Республикамызда ўзбек компьютер лингвистикаси бўйича А.Қ.Пўлатов ва М.Х.Ҳакимов кўп тилли машина таржимаси муҳитидаги энг долзарб муаммолар, расмий грамматиканинг ўрни, расмий грамматикадан алгоритмларга ўтишда зарур бўлган механизм сифатида табиий тилларни математик моделлаштириш масаласи ва бу масалани ҳал қилиш учун яратилган сўз туркумини ҳосил қилишнинг математик моделлари, семантик базасини яратиш ҳамда ўзбек тили ва инглиз тилларининг солиштирма таҳлили бўйича илмий изланишлар олиб борганлар. Ш.А.Назиров туркий тиллар оиласига мансуб бўлган тилларни таржима қилиш учун электрон

луғатни шакллантириш ва маълумотлар базасини ишлаб чиқиш бўйича бир қатор ишларни амалга оширган. Ўзбек тилидаги сўзларни бўғинларга ажратиш бўйича М.Арипов ва уларнинг шогирдлари изланишлар олиб бормоқдалар. Матнни овозли таниб олиш ва талаффузида нуқсони бор болаларга ўзбек тилидаги ҳарфларни тўғри ўқишни ўргатувчи таржимон дастурлар билан М.Мусаев ва шогирдлари шуғулланиб келишмоқда.

Юқорида келтириб ўтилган олимларнинг ишлари ўзбек компьютер лингвистикаси тадқиқоти учун ҳозирги кунда таянч материал сифатида фойдаланиб келинаётган бўлса-да, бироқ ҳозиргача туркий тиллар оиласига мансуб бўлган тиллар бўйича автоматик таржимани яратишга доир ўзбек лингвистикасининг компьютерга йўналтирилган моделлари, алгоритмлари ва дастурий мажмуалари етарли даражада ўрганилмаган.

Диссертация тадқиқотининг диссертация бажарилган олий таълим муассасаси илмий-тадқиқот ишлари билан боғлиқлиги. Диссертация тадқиқоти Тошкент ахборот технологиялари университети илмий-тадқиқот ишлари режасининг №А5-015 «Туркий тиллар учун электрон таржиманинг математик модели, алгоритми ва дастурий таъминоти» (2012-2014), №БВ-Атех-2018-251-сонли «Кўзи ожиз шахслар учун компьютер техникасидан фойдаланиш, матнларни ўқиш ва ёзиш имконини берувчи ўзбек тилига асосланган гапирувчи дастурий таъминот ва овоз синтезаторини ишлаб чиқиш» (2018-2020) мавзуларидаги лойиҳалар доирасида бажарилган.

Тадқиқотнинг мақсади ўзбек тилидан қорақалпоқ тилига таржимани амалга оширишнинг моделлари ва алгоритмлари, икки тилли электрон луғат учун маълумотлар базасини қуриш ҳамда дастурий мажмуасини ишлаб чиқишдан иборат.

Тадқиқотнинг вазифалари:

икки тилли электрон таржимон дастурларни таҳлил қилиш ва туркий тиллар (ўзбек-қорақалпоқ мисолида) электрон луғат учун маълумотлар базасини яратиш;

ўзбек ва қорақалпоқ тилларини лексик, морфологик таҳлил қилиш, мураккаб сўз бирикмаларини ҳамда содда гап тузилмасини ҳосил қилиш, математик моделини ишлаб чиқиш;

ўзбек ва қорақалпоқ тиллари учун мураккаб сўз ясалишининг математик модели асосида самарали алгоритмларни, икки тилли электрон луғат ҳамда таржимон учун маълумотлар базаси асосида сўзларни қидириш ва матнларнинг ўзаро таржимасини аниқлаш алгоритминини ишлаб чиқиш;

туркий тилларга оид электрон таржимон учун ҳисоблаш алгоритмларини модулли таҳлил қилиш асосида дастурлаш жараёнларини формаллаштириш ҳамда дастурий мажмуанинг IDEF0 модели, дастурий модулларнинг функционал тузилмаси ва маълумотлар базасининг инфологик IDEF1x моделини ишлаб чиқиш;

таклиф қилинган модел ва алгоритмлар асосида икки тилли ўзбек-қорақалпоқ ва қорақалпоқ-ўзбек матнлари таржимасини амалга оширадиган дастурий мажмуа ва фойдаланувчи интерфейсини яратиш.

Тадқиқотнинг объекти ўзбек тилидан қорақалпоқ тилига таржима қилувчи электрон таржимонлар, ўзбек ва қорақалпоқ тилининг лингвистик корпуслари ҳисобланади.

Тадқиқотнинг предмети Туркий тиллар учун икки тилли электрон таржиманинг алгоритмлари ва дастурий мажмуасини яратиш технологияларидир.

Тадқиқотнинг усуллари. Тадқиқот мавзусини ёритишда лингвистик ва сонли моделлаштириш, ҳисоблаш алгоритмларини модулли таҳлил қилиш усулларидан, замонавий дастурлаш технологияларидан фойдаланилди.

Тадқиқотнинг илмий янгилиги қуйидагилардан иборат:

ўзбек ва қорақалпоқ тилларининг мантиқий лингвистик моделлари асосида уларнинг электрон луғат маълумотлар базаси яратилган;

ўзбек ва қорақалпоқ тилларининг лексик, морфологик-семантик базаси асосида мураккаб сўз бирикмаларини ҳамда содда гап тузилмасини ҳосил қилишнинг математик модели ва алгоритми ишлаб чиқилган;

ўзбек ва қорақалпоқ тиллари учун икки тилли электрон луғат ҳамда электрон таржимон учун маълумотлар базаси асосида сўзларни қидириш ва матнларнинг ўзаро таржимасини аниқлаш алгоритми ишлаб чиқилган;

ўзбек ва қорақалпоқ тилларида электрон таржима алгоритмларининг модулли таҳлили асосида дастурлаш жараёнларининг функционал модуллари ишлаб чиқилган;

ўзбек ва қорақалпоқ тилларига оид электрон таржима алгоритмларини дастурлаш жараёнларининг функционал модуллари асосида дастурий мажмуа тузилмаси ишлаб чиқилган ҳамда фойдаланувчи интерфейси яратилган.

Тадқиқотнинг амалий натижаси қуйидагилардан иборат:

икки тилли электрон луғат ҳамда таржимон маълумотлар базаси, семантик таҳлил асосида сўзларни қидириш ва матнларнинг ўзаро таржимасини аниқлаш усули ишлаб чиқилган;

икки тилли электрон таржимонни амалга ошириш учун ўзбек ва қорақалпоқ тилларида мураккаб сўзларни ясаш математик модели ишлаб чиқилган;

икки тилли электрон таржимонни амалга ошириш учун ўзбек ва қорақалпоқ тилларида содда гапларни тузиш учун математик модел ишлаб чиқилган;

маълумотлар базасининг инфологик IDEF1x ва билимлар базасининг бошқаришга йўналтирилган объектли модели ишлаб чиқилган;

дастурий мажмуанинг ишлаш тамойили ва зарурий талаблар ишлаб чиқилган.

Тадқиқот натижаларининг ишончлилиги тадқиқот натижаларининг ишончлилиги икки туркий тиллар орасида содда гапларни таржима қилиш махсус дастурий тўпламдан фойдаланган ҳолда ўтказилган тажрибалар натижаларининг ишонччилигини исботлашга асосланган якуний хулосалар билан изоҳланади.

Тадқиқот натижаларининг илмий ва амалий аҳамияти. Тадқиқот натижаларининг илмий аҳамияти икки тилли туркий тиллар ўзбек ва

қорақалпоқ тилларида таржима жараёнини амалга оширишда мураккаб сўзларни ясаш ва содда гапларни тузишнинг математик модел ва сонли алгоритмини яратиш билан изоҳланади. Тадқиқот натижаларининг амалий аҳамияти шундан иборатки, яратилган математик модел ва алгоритмлар мураккаб сўзларни ясаш ва содда гапларни тузишда ишлатиладиган тегишли дастурий мажмуани ишлаб чиқиш билан изоҳланади.

Тадқиқот натижаларининг жорий қилиниши. Диссертация доирасида олиб борилган тадқиқотлар натижасида ишлаб чиқилган алгоритмлар ва яратилган дастурий мажмуа асосида:

Ўзбек ва қорақалпоқ тилларида мураккаб сўзларни ясаш алгоритми ва математик модели асосида ишлаб чиқилган дастурий таъминот «Респект» таржималар бюросида матнларни таржима қилишда юзага келадиган камчиликларни бартараф этишда жорий қилинган (Ахборот технологиялари ва коммуникацияларини ривожлантириш вазирлигининг 2021 йил 19 февралдаги 33-8/1255-сон маълумотномаси). Натижада умумий иш фаолиятининг фойдали иш коэффициенти 90%га ошишига имкон берди.

Ўзбек ва қорақалпоқ тилларида содда гапларни тузиш алгоритми ва математик модели асосида ишлаб чиқилган дастурий мажмуа «FARMA PRINT NUKUS» МЧЖга жорий қилинган (Ахборот технологиялари ва коммуникацияларини ривожлантириш вазирлигининг 2021 йил 19 февралдаги 33-8/1255-сон маълумотномаси). Натижада умумий иш фаолиятининг ҳажмини вақтга нисбатан 1-2 баробар қисқартиришга имкон берди.

Ўзбек тилида мураккаб сўзларни ҳосил қилиш алгоритми ва математик модели асосида ишлаб чиқилган дастурий мажмуа «ROBOTICS LAB FUTURE TECHNOLOGY COMPANY» МЧЖга жорий қилинган (Ахборот технологиялари ва коммуникацияларини ривожлантириш вазирлигининг 2021 йил 19 февралдаги 33-8/1255-сон маълумотномаси). Натижада умумий иш фаолиятида сўзларни қайта ишлаш тезлигининг 50% га ошишига имкон берди.

Тадқиқот натижаларининг апробацияси. Ушбу тадқиқот натижалари жами 16 та, жумладан, 7 та халқаро ва 9 та республика илмий-амалий анжуманларида муҳокамадан ўтказилган.

Тадқиқот натижаларининг эълон қилинганлиги. Тадқиқотнинг асосий натижалари 4 та илмий ишларда эълон қилинган, улардан 3 таси Ўзбекистон Республикаси Олий аттестация комиссияси томонидан докторлик диссертацияларининг асосий илмий натижаларини эълон қилиш учун тавсия қилинган журналларда, жумладан, 1 таси хорижий журналларда ва 2 таси республика журналларида нашр қилинган, Scopus базасига кирувчи журналда 1 та мақола нашр қилинган ҳамда 2 та ЭҲМ учун дастурларни расмий рўйхатдан олинганлиги тўғрисида гувоҳнома олинган.

Диссертациянинг тузилиши ва ҳажми. Диссертация таркиби кириш, тўртта боб, умумий хулоса, фойдаланилган адабиётлар рўйхати ва иловалардан иборат. Диссертациянинг ҳажми 128 бетни ташкил этади.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Кириш қисмида диссертация мавзусининг долзарблиги ва зарурияти асосланган, тадқиқотнинг Ўзбекистон Республикаси фан ва технологиялар тараққиётининг устувор йўналишларига мослиги кўрсатилган, тадқиқотнинг мақсад ва вазифалари белгилаб олинган ҳамда тадқиқот объекти ва предмети аниқланган, олинган натижаларнинг ишончлилиги асослаб берилган, уларнинг назарий ва амалий аҳамияти, тадқиқот натижаларини амалда жорий қилиш ҳолати, нашр этилган ишлар ва диссертация тузилиши бўйича маълумотлар келтирилган.

Диссертациянинг «**Электрон таржимонлар ва компьютер лингвистикасининг ривожланиш босқичлари**» деб номланган биринчи бобининг биринчи параграфида электрон таржимоннинг ҳозирги ҳолати ва ривожланиш истиқболлари келтирилган. Иккинчи параграфида ҳозирги кунда мавжуд электрон таржимонларнинг таҳлили, уларнинг авфзал ва камчиликлари таҳлил қилинган. Учинчи параграфида туркий тиллар учун электрон таржимонларни яратиш муаммолари ўрганилган.

Бир туркий тилдан иккинчисига электрон таржима қилиш механизмини ёритишда қуйидагилар эътиборга олинади:

1. Аксарият туркий тилларда кирилл ва лотин алфавитлари ишлатилади. Ҳар бир туркий тилларда ўзига хос товушларни ифодаловчи ҳарфлар ҳам мавжуд. Масалан, ўзбек тилида кирил алифбосидаги ҳарфлардан ташқари яна тўртта ҳарф – ў, қ, ғ, ҳ ҳарфлари мавжуд. Лотин алифбосида бу ҳарфлар фарқли фонетик белгилар билан белгиланади. Худди шундай ҳарфлар бошқа туркий тилларда ҳам мавжуд. Шунини таъкидлаш жоизки, ҳатто бир туркий тилларда бир нечта алифболар ишлатилиши мумкин. Масалан, Ўзбекистонда ўзбеклар лотин графикасидан фойдаланишса, Афғонистондаги ўзбеклар араб графикасидан фойдаланишади. Шунинг учун ҳам туркий тилларнинг барча алифболарини ўзаро қўллаш лозим бўлади.

2. Ҳар бир туркий тил учун олд ва ўзакдан кейинги қўшимчаларни, яъни префикс ва суффикслар ҳамда сўз ўзаклари бўйича маълумотлар базаларини ташкил этиш керак. Худди шундай базалар келишиклар, шахслар учун ҳам яратилади. Мазкур базаларни кирилчада ва лотинчада алоҳида-алоҳида ташкил этиш лозим.

3. Ҳар бир туркий тил ва гуруҳ ичидаги тиллар учун тилларнинг формал моделини яратиш.

4. Туркий тилларнинг электрон луғатини яратиш.

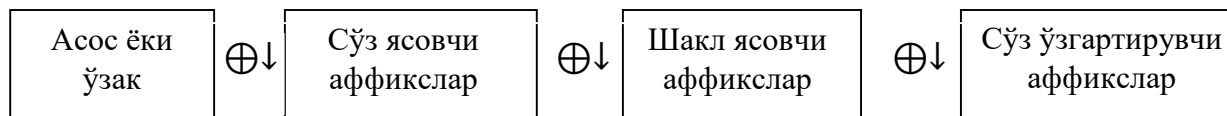
5. Туркий тиллардаги синоним бирликларни қайта фразалаш учун маълумотлар базасини яратиш.

6. Электрон таржима учун дастурий таъминот функционал тўлдирмасини яратиш.

7. Электрон таржимани бошқарувчи моделини яратиш.

8. Дастурий таъминотни тестлаш ва кўриқдан ўтказиш, верификация ва волидацияларини таъминлаш.

Диссертациянинг «Туркий тиллар учун мураккаб сўзларни ҳосил қилишнинг математик модели» номли иккинчи бобида ўзбек ва қорақалпоқ тилларида мураккаб сўзларни ҳосил қилувчи математик моделлар ишлаб чиқилган бўлиб, мураккаб сўзларни яшаш қуйидаги схема бўйича аффиксларни ўзакка қўшиш орқали яратилади.



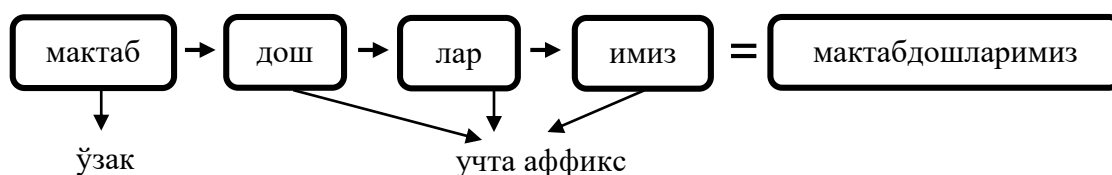
1-расм. Мураккаб сўз ясовчи схема

1-расмда кўрсатилган схема асосида биз ўзбек тилида мураккаб сўз яшашнинг математик моделини ишлаб чиқдик.

$$C = \sum_{i=1}^n x_i \oplus \downarrow \sum_{j=1}^m a_j \oplus \downarrow \sum_{k=1}^l f_k \oplus \downarrow \sum_{r=1}^q p_r$$

Бу ерда C – мураккаб сўз, $\sum_{i=1}^n x_i$ – ҳар бир сўз x тўплам ҳисобланадиган сўз асоси, $\sum_{j=1}^m a_j$ – сўз ясовчи аффикслар, $\sum_{k=1}^l f_k$ – шакл ясовчи аффикслар, $\sum_{r=1}^q p_r$ – сўз ўзгартирувчи аффикслар, бу ерда i, j, k ва r – аффикслар ва сўз узунлиги. Бу ерда \oplus – бирлаштириш амалини, \downarrow – «боғлаш» (бирлаштириш ёки «боғланмаслик» (бирлашмаслик) амалларини билдирувчи белгилар.

Юқоридаги схема ва мураккаб сўзларни яратиш учун математик модел асосида мисол келтирамиз (2-расм).



2-расм. Ўзбек тилида сўз ясалиш схемасидан фойдаланиш мисоли

1. «мактаб» - сўз асоси, ўзак
2. «дош» - сўз ясовчи аффикс,
3. «лар» -шакл ясовчи аффикс,
4. «имиз» -сўз ўзгартирувчи аффикс.

Мисолдан кўриниб турибдики, сўз ўзбек тилининг учта сўз-қўшимчасини сўзнинг илдизига қўшиш орқали ҳосил бўлади.

Юқоридаги математик модел асосида биз ўзбек тилида мураккаб сўзларни яратиш кетма-кетлигини кўриб чиқамиз.

$$C_1 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_2 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_3 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_4 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

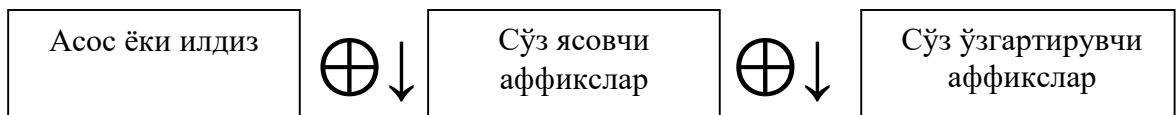
$$C_5 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_6 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_7 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{бу ерда } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

Бу ерда $C_1, C_2, C_3, C_4, C_5, C_6$ ва C_7 – сўз асоси, $I(z)$ – битта предикат, ҳар бир z сўзи икки ёки ундан ортиқ сўз бирикмаси сифатида ифодаланади $y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1}$, бу ерда i – аффикс узунлиги.

Қорақалпоқ тили ҳам аглютинатив тиллар гуруҳига киради, мураккаб сўзларни ҳосил қилиш ўзбек тили каби амалга ошади (3-расм).



3-расм. Қорақалпоқ тилида мураккаб сўзларни ҳосил қилиш схемаси

Диссертация учинчи «**Икки тилли туркий тиллар учун гапларни тузишда семантик модел алгоритмларини қўллаш кетма-кетлиги**» бобида ўзбек ва қорақалпоқ тилларида содда ва мураккаб гапларни тузиш схемаси, математик модели ва алгоритми ишлаб чиқилган.

Маълумки, ўзбек тилида гапда сўзларнинг одатдаги тартиби мавжуд. Бу тартибга кўра гапда эга олдин, кесим эса кейин жойлашади, аниқловчи аниқланмишдан, ҳол ва тўлдирувчи кесимдан олдин туради, демак, кесим одатда гапнинг охирида келади.

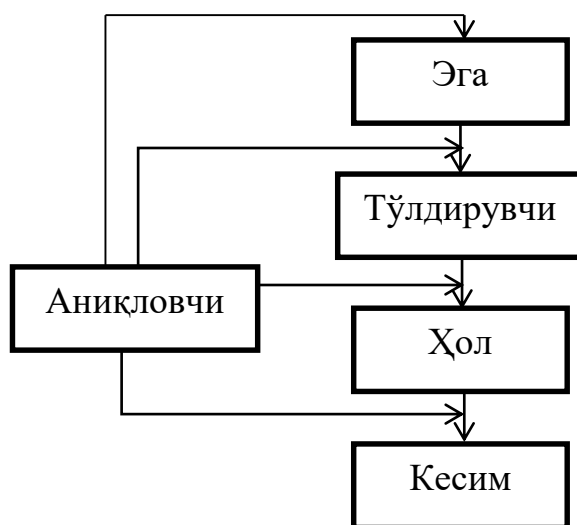
Аммо муайян услубий маънони ифодалаш мақсади билан гапдаги сўзларнинг одатдаги тартиби ўзгартирилиши мумкин, бундай ўзгарган сўз тартиби инверсия деб ҳам юритилади.

Илмий ва расмий услубларда гапда сўзларнинг одатдаги тартиби, асосан катъий сақланади. Лекин сўзлашув, бадий ва публицистик услубларда бу тартиб ўзгариши, яъни инверсион тартиб қўлланиши мумкин.

Ўзбек тилида содда ва мураккаб гаплар қуйидаги (4 ва 5-расмлар) кетма-кетликларда амалга оширилади:



4-расм. Ўзбек тилида содда гапларни тузиш схемаси



5-расм. Ўзбек тилида мураккаб гапларни тузиш схемаси

Юқоридаги схемалар асосида ўзбек тилида содда ва мураккаб гапларни ҳосил қилишнинг математик моделлари ишлаб чиқилди.

$$G_{\text{сод}} = A(x_i, \dots, x_n) \cup B(x_i, \dots, x_n)$$

$$G_{\text{мур}} = A(x_i, \dots, x_n) \cup D(x_i, \dots, x_n) \cup S(x_i, \dots, x_n) \cup K(x_i, \dots, x_n) \cup B(x_i, \dots, x_n)$$

Бу ерда $G_{\text{сод}}$ – содда гап, $G_{\text{мур}}$ – мураккаб гап, $A(x_i, \dots, x_n)$ – эга, $D(x_i, \dots, x_n)$ – тўлдирувчи, $S(x_i, \dots, x_n)$ - ҳол, $K(x_i, \dots, x_n)$ – аниқловчи, $B(x_i, \dots, x_n)$ – кесим.

$$G_1 = \left\{ \begin{array}{l} A(x_i, \dots, x_n) \\ B(x_i, \dots, x_n) \rightarrow f_{k(\Phi)} \oplus P_{r(\text{от})} \oplus f_{k(\Phi)} \oplus a_{j(\text{от})} \oplus P_{r(\Phi)} \oplus f_{k(\Phi)} \end{array} \right.$$

$$G_2 = \begin{cases} A(x_i, \dots, x_n) \\ K(x_i, \dots, x_n) \rightarrow a_{j(\text{от})} \oplus P_{r(\text{си}\phi)} \oplus P_{r(\phi)} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

$$G_3 = \begin{cases} A(x_i, \dots, x_n) \\ S(x_i, \dots, x_n) \rightarrow P_{r(\text{от})} \oplus P_{r(\text{рав})} \oplus P_{r(\phi)} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

$$G_4 = \begin{cases} A(x_i, \dots, x_n) \\ K(x_i, \dots, x_n) \rightarrow a_{j(\text{от})} \oplus P_{r(\text{си}\phi)} \oplus P_{r(\phi)} \\ D(x_i, \dots, x_n) \rightarrow f_{k(\text{рав})} \oplus P_{r(\text{от})} \oplus P_{r(\phi)} \oplus P_{r(\text{рав})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

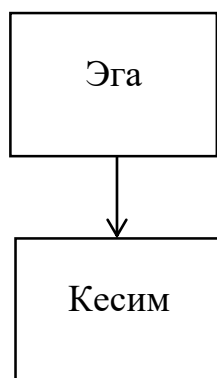
$$G_5 = \begin{cases} A(x_i, \dots, x_n) \\ D(x_i, \dots, x_n) \rightarrow f_{k(\text{рав})} \oplus P_{r(\text{от})} \oplus P_{r(\phi)} \oplus P_{r(\text{рав})} \\ S(x_i, \dots, x_n) \rightarrow P_{r(\text{от})} \oplus P_{r(\text{рав})} \oplus P_{r(\phi)} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

$$G_6 = \begin{cases} K(x_i, \dots, x_n) \rightarrow a_{j(\text{от})} \oplus P_{r(\text{си}\phi)} \oplus P_{r(\phi)} \\ A(x_i, \dots, x_n) \\ D(x_i, \dots, x_n) \rightarrow f_{k(\text{рав})} \oplus P_{r(\text{от})} \oplus P_{r(\phi)} \oplus P_{r(\text{рав})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

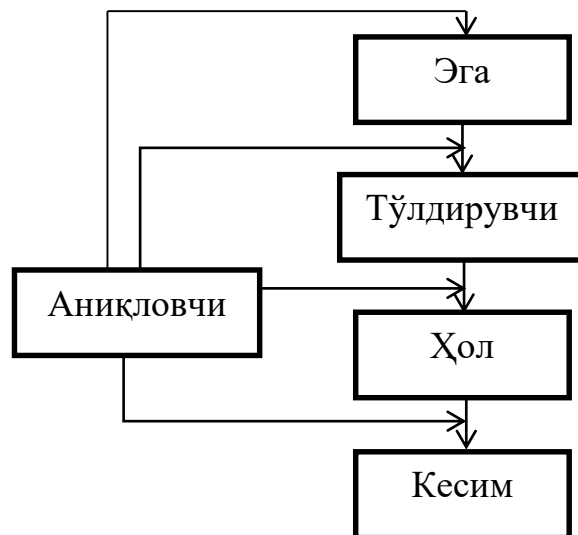
$$G_7 = \begin{cases} D(x_i, \dots, x_n) \rightarrow f_{k(\text{рав})} \oplus P_{r(\text{от})} \oplus P_{r(\phi)} \oplus P_{r(\text{рав})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

$$G_8 = \{B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)}\}$$

Қорақалпоқ тилида ҳам содда ва мураккаб гапларни тузиш ўзбек тилидаги каби амалга оширилади (6, 7-расм).



6-расм. Қорақалпоқ тилида содда гапларни тузиш схемаси



7-расм. Қорақалпоқ тилида мураккаб гапларни тузиш схемаси

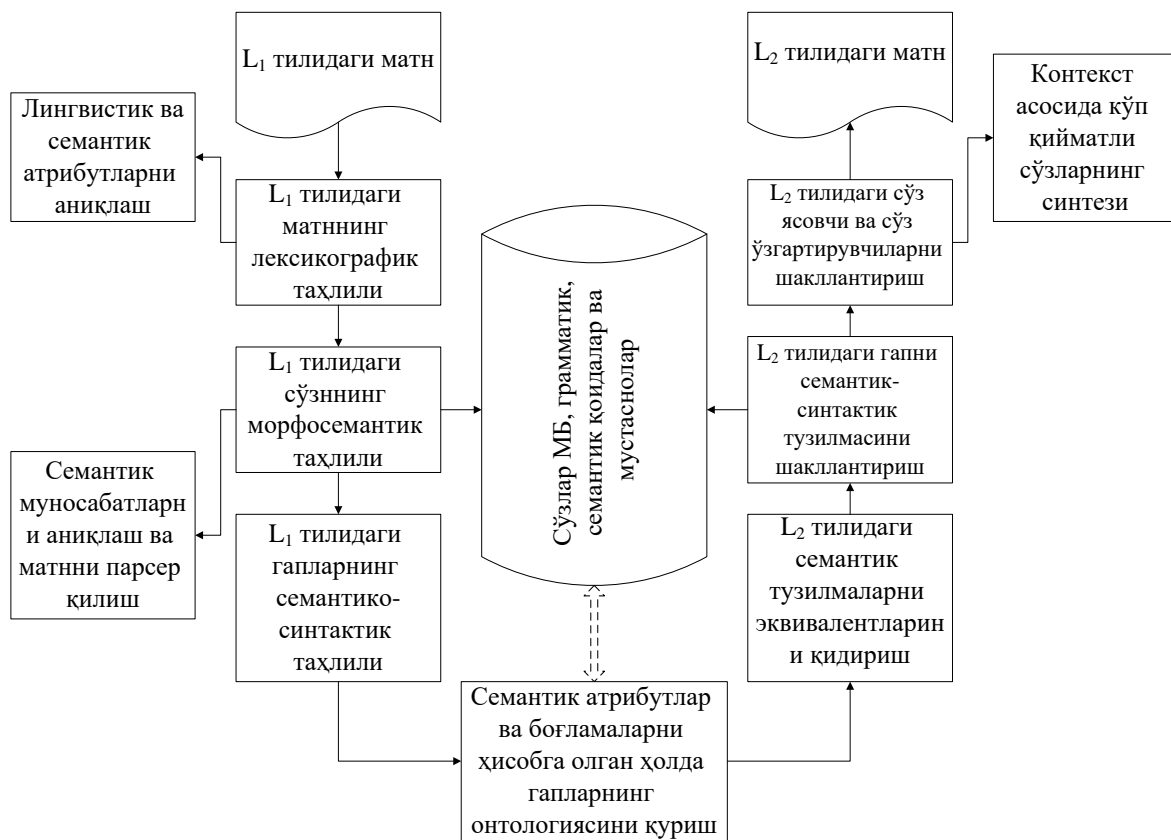
Диссертациянинг «**Икки тилли туркий тиллар учун электрон луғат ва таржимоннинг дастурий мажмуасини яратиш**» номли тўртинчи бобида электрон таржимон маълумотлар базасининг инфологик модели, электрон луғат ва таржимоннинг функционал модули ҳамда дастурий таъминотдан фойдаланиш йўриқномаси ишлаб чиқилган.

Кўриб чиқилаётган предмет соҳани объектлари орасида қуйидаги типдаги боғланишлар мавжуд:

1. 1:1 – объектлар билан атрибутлар орасидаги боғланиш.

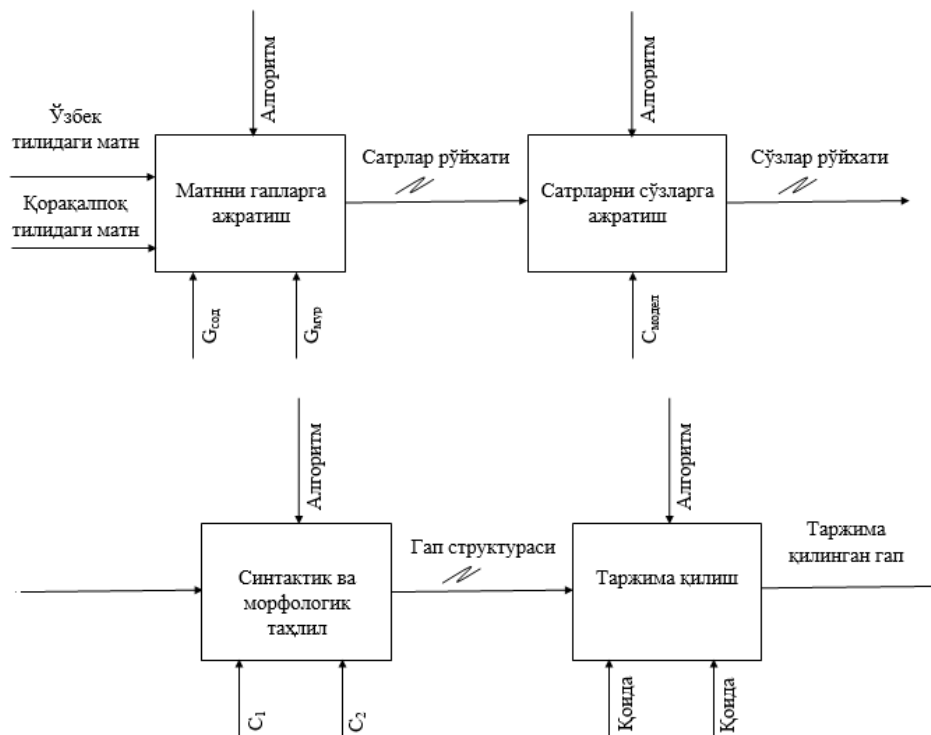
Шундай қилиб, кўрилаётган масалада асосан объектлар аниқланади ва улар орасидаги боғланиш топилади, синфларга ажратилади (8-расм).

1. Таржима қилинадиган матн киритилади.
2. Таржима қилинаётган тилдаги матннинг лексикографик таҳлили: Лингвистик ва семантик атрибутларни аниқлаш.
3. Таржима қилинаётган тилдаги сўзнинг морфосемантик таҳлили: Семантик муносабатларни аниқлаш ва матнни парсер қилиш.
4. Таржима қилинаётган тилдаги гапларнинг семантико-синтактик таҳлили.
5. Семантик атрибутлар ва боғламаларни ҳисобга олган ҳолда гапларнинг онтологиясини қуриш.
6. Таржима қилинадиган тилдаги семантик тузилмаларни эквивалентларини қидириш.
7. Таржима қилинадиган тилдаги гапни семантик-синтактик тузилмасини шакллантириш.
8. Таржима қилинадиган тилдаги сўз ясовчи ва сўз ўзгартирувчиларни шакллантириш: Контекст асосида кўп қийматли сўзларнинг синтези.
9. Таржима қилинган матн.

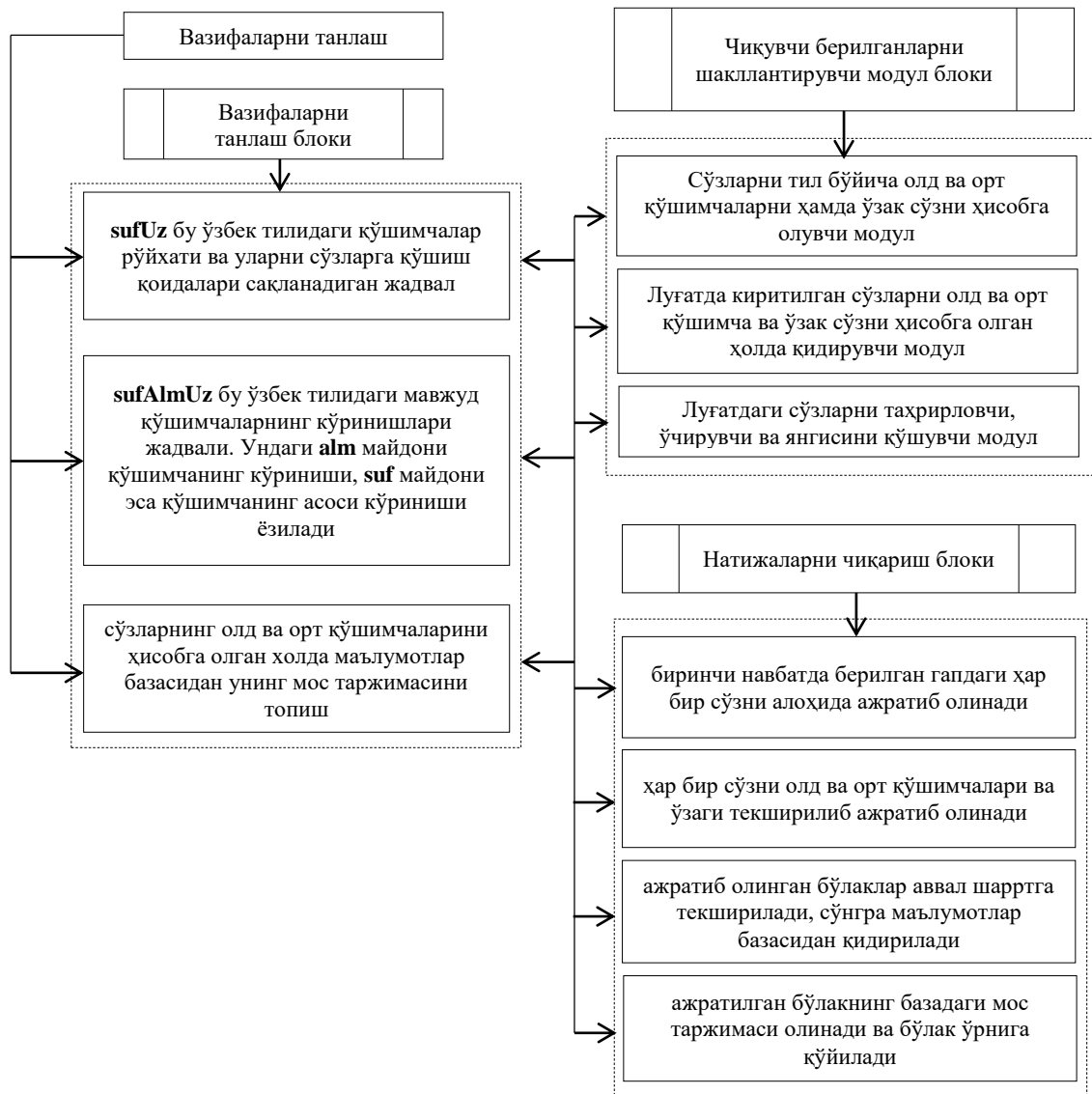


8-расм. Ўзбек тилидан қорақалпоқ тилига машина таржимасини семантик таҳлил ва генератор тизимининг схемаси

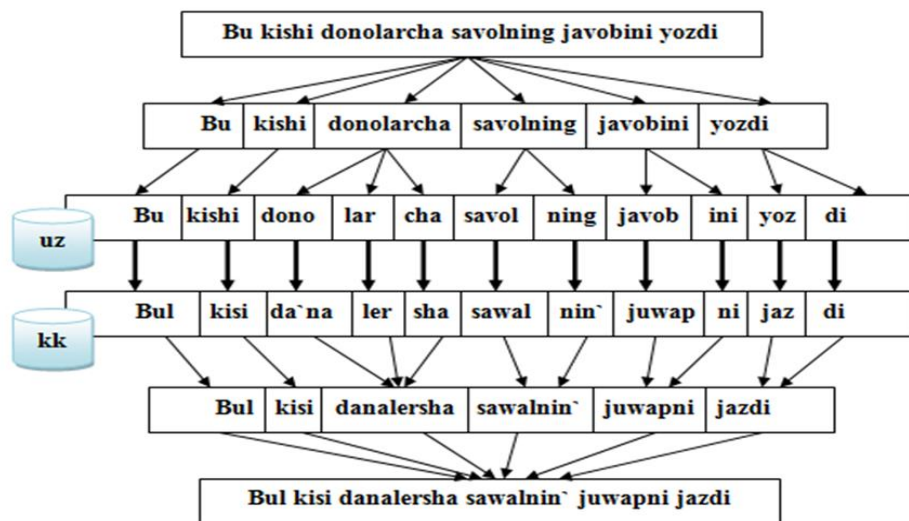
Электрон таржимоннинг IDEF0 схемаси 9-расмда тасвирланган.



9-расм. Электрон таржимоннинг IDEF0 схемаси

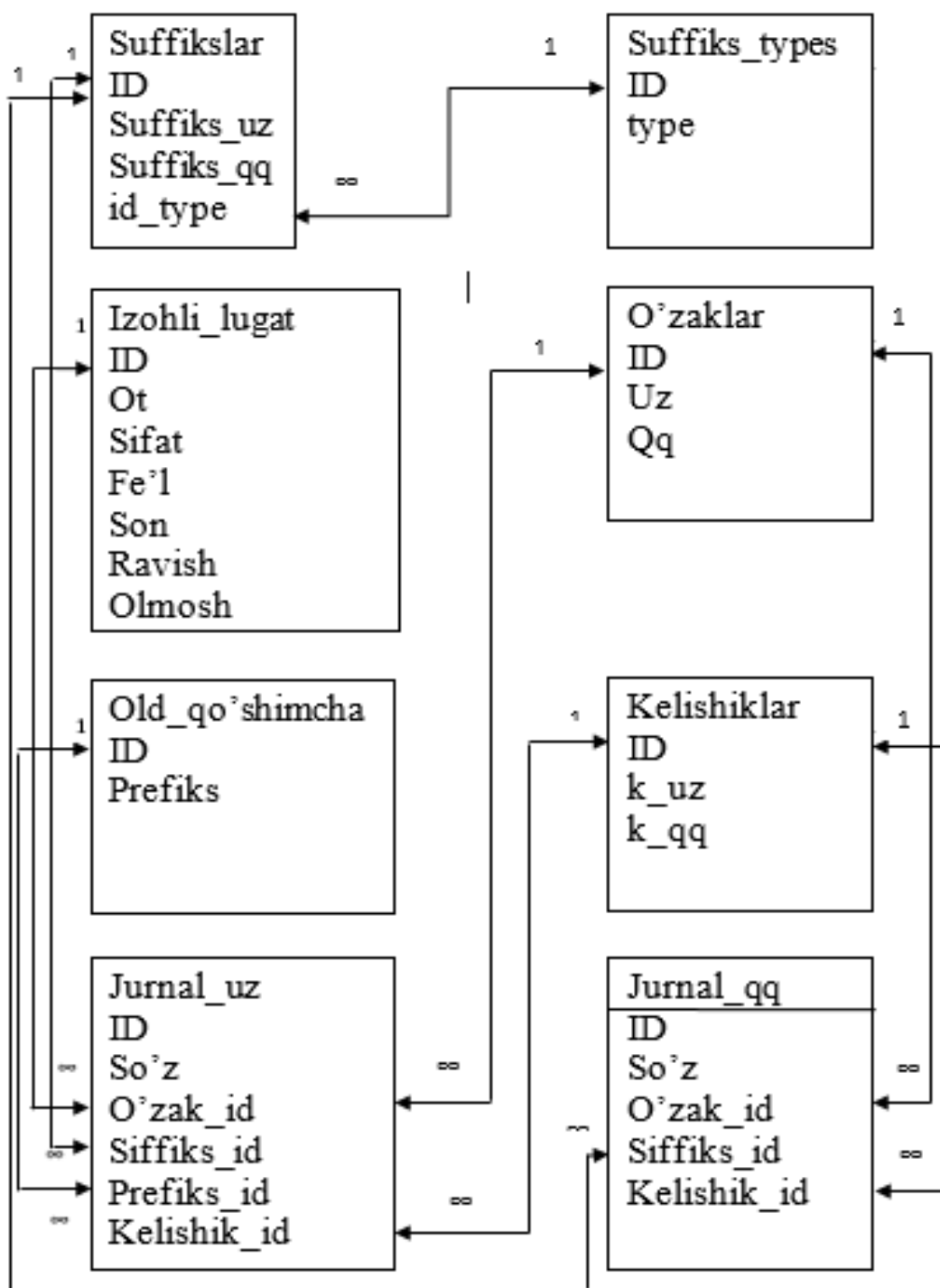


10-расм. Дастурий таъминот ишлашининг функционал схемаси



11-расм. Семантик таҳлил асосида таржимани амалга ошириш алгоритми

Ажратиб олинган префикс, ўзак ва суффикс мос равишда маълумотлар омборидан текширилади. Текширилгандан сўнг янги сўз хосил қилинади ва чоп этилади. Агарда жадвалдан сўз топилмаса уни шундоқ таржима матни сифатида чоп этилади (10-12-расм).



12-расм. Электрон таржимоннинг IDEF1x схемаси

Агарда таржима қилинишда фойдаланилаётган сўзлар маълумотлар базасида учрамаса фойдаланувчига огоҳлантириш беради ёки шу сўзни ўзини чиқаради.

ХУЛОСА

«Туркий тиллар учун электрон таржиманинг моделлари ва алгоритмлари (ўзбек тилидан қорақалпоқ тилига таржима мисолида)» мавзусидаги диссертация иши бўйича олиб борилган тадқиқотлар натижасида қуйидаги хулосалар тақдим этилади:

1. Ўзбек ва қорақалпоқ тилларининг лексик, морфологик таҳлил қилиш ва мураккаб сўз бирикмаларини яшаш математик модели ишлаб чиқилди, бу ўзбек ва қорақалпоқ тилларида содда ва мураккаб сўзларни ҳосил қилишга имкон берди, ушбу математик модел асосида ўзбек ва қорақалпоқ тиллари учун электрон луғат маълумотлар базаси ва билимлар базаси ишлаб чиқилди;
2. Ўзбек ва қорақалпоқ тиллари учун мураккаб сўз яшаш математик модели асосида самарали алгоритми яратилган бўлиб, ушбу алгоритм ўзбек ва қорақалпоқ тилларида сўзларни яшаш кетма-кетлигини келтиришга имкон берди, ушбу алгоритм асосида туркий тиллар учун электрон таржимон яратиш дастурий воситаси ишлаб чиқилди;
3. Икки тилли электрон луғат ҳамда таржимон маълумотлар базаси, асосида сўзларни қидириш ва матнларни ўзаро таржимасини аниқлаш алгоритмлари ишлаб чиқилган бўлиб, семантик таҳлилдан фойдаланган ҳолда ўзбек тилидан қорақалпоқ тилига таржима жараёни амалга оширилди, ушбу семантик таҳлил асосида сўзларнинг ечимлар дарахти қурилди;
4. Туркий тиллар электрон таржимон дастурий мажмуини IDEF0 модели, дастурий модулларнинг функционал тузилмаси ва маълумотлар базасини инфологик IDEF1x модели ва билимлар базасининг объектга йўналтирилган моделларини ишлаб чиқиш масаласи ечилган. Туркий тиллар учун электрон таржимон дастурий мажмуаси модуллари ва унинг таркибий қисмлари, маълумотлар базасининг тузилмаси, билимлар базасининг тузилмаси ва уларда бажариладиган ишлар белгилаб олинган;
5. Таклиф қилинган модел ва алгоритмлар асосида, икки тилли ўзбек-қорақалпоқ ва қорақалпоқ-ўзбек матнларини таржимасини амалга оширадиган дастурий мажмуа ва қулай фойдаланувчи интерфейси яратилди. Дастурий мажмуа фойдаланувчилар учун таржимани тез ва сифатли амалга оширишга хизмат қилади;
6. Яратилган дастурий восита «Респект» таржимонлик бюроси, «FARMA PRINT NUKUS» МЧЖ ва «ROBOTICS LAB FUTURE TECHNOLOGY COMPANY» МЧЖ умумий иш фаолиятининг ҳажмини вақтга нисбатан 1-2 баробар қисқартиришга имкон берди.

**НАУЧНЫЙ СОВЕТ DSc.13/30.12.2019.Т.07.01 ПО ПРИСУЖДЕНИЮ
УЧЕНЫХ СТЕПЕНЕЙ ПРИ ТАШКЕНТСКОМ УНИВЕРСИТЕТИ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

**ТАШКЕНТСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ**

АБИДОВА ШАХНОЗА БАХОДИРОВА

**МОДЕЛИ И АЛГОРИТМЫ ЭЛЕКТРОННОГО ПЕРЕВОДА ДЛЯ
ТЮРКСКИХ ЯЗЫКОВ (НА ПРИМЕРЕ ПЕРЕВОДА С УЗБЕКСКОГО
ЯЗЫКА НА КАРАКАЛПАКСКИЙ ЯЗЫК)**

05.01.04 – Математическое и программное обеспечение вычислительных
машин, комплексов и компьютерных сетей

**АВТОРЕФЕРАТ ДИССЕРТАЦИИ
ДОКТОРА ФИЛОСОФИИ (PhD) ПО ТЕХНИЧЕСКИМ НАУКАМ**

Ташкент – 2021

Тема диссертации доктора философии (PhD) по техническим наукам зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за В2021.1.PhD/T2080.

Диссертация выполнена в Ташкентском университете информационных технологий имени Мухаммад-ал Хоразми.

Автореферат диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-странице научного совета (www.tuit.uz) и в Информационно-образовательном портале «Ziyonet» (www.ziyonet.uz).

Научный руководитель: **Нуралиев Фахриддин Муродиллаевич**
доктор технических наук, доцент

Официальные оппоненты: **Мухамедиева Дилноз Тулкуновна**
доктор технических наук, профессор

Норов Абдисаит Мурадович
доктор философии технических наук (PhD)

Ведущая организация: Ташкентский государственный технический университет имени Ислама Каримова

Защита диссертации состоится «19» июня 2021 г. в 9:00 часов на заседании научного совета DSc.13/30.12.2019.T.07.01 при Ташкентском университете информационных технологий имени Мухаммад ал-Хоразми. (Адрес: 100202, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-43, факс: (99871) 238-65-52, e-mail: tuit@tuit.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского университета информационных технологий имени Мухаммад ал-Хоразми (регистрационный номер № 109). Адрес: 100202, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-65-44.

Автореферат диссертации разослан «15» июня 2021 г.
(протокол рассылки № 17 «10» июня 2021 г.).



Р.Х. Хамдамов
Председатель научного совета
по присуждению ученых степеней, д.т.н., профессор

Э.Ш. Назирова
Ученый секретарь научного совета
по присуждению ученых степеней, д.т.н., доцент

М.А. Рахматуллаев
Председатель научного семинара при научном
совете по присуждению ученых степеней,
д.т.н., профессор

ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))

Актуальность и востребованность темы диссертации. В мировой лингвистике, начиная с двадцатого века, растет интерес к созданию, регулированию и научному изучению системы терминов в области компьютерной лингвистики. В связи с огромными изменениями в социально-политической, экономической и духовной жизни общества одной из наиболее актуальных проблем является машинное обучение терминов, принадлежащих к тюркской языковой семье, и изучение компьютерной лингвистики, внедрение процессов компьютерного моделирования. Действительно, создание компьютерно-ориентированных моделей естественного языка и решение ряда задач на основе этих моделей, автоматический морфологический, синтаксический и семантический анализ естественных языков, а также выявление их особенностей важны в компьютерной лингвистике. В развитых странах компьютерной лингвистики, включая Россию, Японию, Германию, США, Китай, Англию, Корею, Францию, Австралию и другие развитые страны, важно разработать методы, модели и программные технологии, используемые в процессе перевода с одного языка на другой.

В мире одной из актуальных проблем алгоритмизации является разработка алгоритмических методов доказательства теорем на основе современных технологий программирования, основанных на вычислительной математике, базу классических теорий, а также необходимых разделов математического и функционального анализа. Алгоритмическая основа систем вопросов-ответов строится путем манипулирования данными на естественном языке. Это также включает проблему алгоритмического программирования, связанную с переключением с одного языка на другой и построением алгоритмических основ для построения интеллектуальных систем, среди прочего. В алгоритмических банках важно заполнить информационные части и разработать программное обеспечение для их операционных частей. Актуальной задачей является создание компьютерной лингвистики узбекского языка, относящегося к семейству тюркских языков, и оптимизация работы электронных переводчиков. Это позволяет создавать компьютерно-ориентированные модели узбекского языка, то есть особое внимание уделяется изучению узбекского языка, оценке знаний, тестированию текстов на узбекском языке, голосовому вводу в компьютер, редактированию текстов.

В республике проводятся научные исследования по созданию двуязычных электронных переводчиков, созданию базы данных полевых слов, созданию корпуса их аннотированных текстов. Примеры включают лингвистико-логическую модель узбекского языка, формирование их компьютерной лингвистики, формирование национального корпуса и базы данных узбекского языка, разработку электронных словарей, моделей автоматического перевода между тюркским и другими языками и многое другое. Стратегия действий по дальнейшему развитию Республики

Узбекистан на 2017-2021 годы включает «...развитие национального содержания, обучение на государственном языке, научно-образовательное, совершенствование механизмов создания и продвижения современных информационных ресурсов, мультимедийных продуктов в соответствии с потребностями молодежи, внедрение информационно-коммуникационных технологий». Для решения этих задач необходимо создание компьютерно-ориентированных моделей узбекского языкознания. Из-за отсутствия исследований в компьютерной лингвистике по машинному переводу на узбекский язык необходимо создать лингвистическую базу данных узбекского языка для программного перевода с узбекского на иностранный язык и наоборот, и одним из важных вопросов является создание алгоритмов и программного обеспечения для их практического использования.

В указе Президента Республики Узбекистан от 21 октября 2019 года № ПФ-5850 «О мерах по кардинальному повышению престижа и статуса узбекского языка как государственный язык» говорится, что «...обеспечение того, чтобы государственный язык занял достойное место в глобальной информационной сети информационно-коммуникационных технологий, в частности, Интернет, создание компьютерных программ на узбекском языке». Настоящее диссертационное исследование послужит делу выполнения задач, установленных нормативными правовыми актами, относящимися к этой деятельности.

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий Республики IV. «Информатизация и развитие информационно-коммуникационных технологий».

Степень изученности проблемы. В мире было проведено много исследований по разработке модели и алгоритмов электронных переводчиков. В диссертации изучены и проведены сравнительные исследования произведений основателя электронного переводчика У. Уивера, Р. Риченса, А.А. Ляпунова, Г.Г. Белоногова, В.Г. Васильева, Л.А. Гращенко, Ю.Н. Марчука и других. Параллельно исследования в этой области проводились в лаборатории инженерной лингвистики Ленинградского государственного педагогического института имени А.И. Герцена.

В Республике Узбекистан по компьютерной лингвистике А.К. Пулатовым и М.Х.Хакимовым проведены научные исследования по наиболее актуальным проблемам многоязычной среды машинного перевода, роли официальных текстов, вопросам математического моделирования естественных языков как необходимого при переходе от официальных текстов к алгоритмам, а также математическим моделям формирования лексики, созданным для решения этой проблемы, созданию семантической базы и сравнительному анализу языковых средств. Ш.А.Назирова провел ряд исследований по формированию электронного словаря и разработке базы данных для перевода языков, принадлежащих к тюркской языковой семье. Исследованием деления слов в узбекском языке на слоги занимаются М.

Арипов и его ученики. М.Мусаев и его ученики занимаются звуковым распознаванием текста, переводческими программами, которые учат детей с дефектами речи правильно читать буквы узбекского языка.

Хотя работы вышеназванных ученых в настоящее время используются в качестве базового материала для исследований узбекской компьютерной лингвистики, но до сих пор компьютерно-ориентированные модели, алгоритмы и программные комплексы узбекской лингвистики по созданию электронного перевода на языки, относящиеся к тюркской семье языков, недостаточно изучены.

Связь диссертационного исследования с планами научно-исследовательских работ высшего образовательного учреждения, где выполнено диссертация. Диссертационное исследование выполнено в рамках проектов №А5-015 «Математическая модель, алгоритм и программное обеспечение электронного перевода на тюркские языки» (2012-2014), № БВ-Атекс-2018-251 «Разработка речевого программного обеспечения и синтезатора голоса на основе узбекского языка, позволяющего использовать компьютерные технологии, чтение и написание текстов для слабовидящих лиц» (2018-2020 годы) согласно научно-исследовательскому плану Ташкентского университета информационных технологий.

Цель исследования построение моделей и алгоритмов перевода с узбекского языка на каракалпакский, базы данных двуязычного электронного словаря и разработка программного комплекса.

Задачи исследования:

анализ двуязычных программ электронного переводчика и создание электронной словарной базы тюркских языков (на примере узбекско-каракалпакского);

лексико-морфологический анализ узбекского и каракалпакского языков, формирование сложных фраз и простой структуры предложений, разработка математической модели;

разработка эффективных алгоритмов на основе математической модели сложного словообразования для узбекского и каракалпакского языков, поиска слов на основе двуязычного электронного словаря и базы данных переводчиков и алгоритма определения взаимного перевода текстов;

формализация процессов программирования на основе модульного анализа вычислительных алгоритмов электронного перевода тюркских языков, а также разработка модели программного комплекса IDEF0, функциональной структуры программных модулей и инфологической модели базы данных IDEF1x;

создание программного комплекса и пользовательского интерфейса, основанного на предложенных моделях и алгоритмах, который переводит узбекско-каракалпакские и каракалпакско-узбекские тексты на два языка.

Объектом исследования являются электронные узбекско-каракалпакские переводчики, лингвистический корпус узбекского и каракалпакского языков.

Предметом исследования являются алгоритмы и технологии разработки программного обеспечения для двуязычного электронного перевода на тюркские языки.

Методы исследования. При освещении темы исследования использовались лингвистическое численное моделирование, методы модульного анализа вычислительных алгоритмов, современные технологии программирования.

Научная новизна исследования заключается в следующем:

на основе лингвистических моделей узбекско-каракалпакских языков создана электронная словарная база тюркских языков;

на основе лексико-морфологического анализа узбекского и каракалпакского языков разработана математическая модель и алгоритм формирования сложной лексики и простой структуры предложения;

разработан двуязычный электронный словарь узбекского и каракалпакского языков, а также алгоритм интерпретатора для поиска слов и определения взаимного перевода текстов на основе данных и базы знаний;

разработаны функциональные модули процессов программирования на основе модульного анализа алгоритмов электронного перевода на узбекский и каракалпакский языки;

на основе функциональных модулей процессов программирования алгоритмов электронного перевода на узбекский и каракалпакский языки была разработана структура программного комплекса и создан пользовательский интерфейс.

Практические результаты исследования заключается в следующем:

на основе двуязычного электронного словаря и базы данных переводчиков разработан семантический анализ, методика поиска слов и определения взаимного перевода текстов;

для реализации двуязычного электронного переводчика разработана математическая модель построения сложных слов в узбекском и каракалпакском языках;

для реализации двуязычного электронного переводчика была разработана математическая модель составления простых предложений на узбекском и каракалпакском языках;

разработана объектная модель, ориентированная на управление инфологической базой данных IDEF1x и базой знаний;

разработан принцип работы программного комплекса и необходимые требования.

Достоверность результатов исследования достоверность результатов исследования объясняется окончательными выводами, основанными на доказательстве достоверности результатов экспериментов, проведенных с использованием специального программного комплекса для перевода простых предложений между двумя тюркскими языками.

Научная и практическая значимость результатов исследования. Научная значимость результатов исследования объясняется тем, что двуязычные тюркские языки составляют сложные слова при осуществлении

процесса перевода на узбекский и каракалпакский языки и создают математическую модель и конечный алгоритм построения простых предложений.

Практическая значимость результатов исследования заключается в том, что созданные математические модели и алгоритмы объясняются разработкой соответствующего программного комплекса, используемого для создания сложных слов и простых предложений.

Внедрение результатов исследования. На основе алгоритмов и программного обеспечения, разработанных в результате исследований, проведенных в рамках диссертации:

программное обеспечение, разработанное на основе алгоритмов и математических моделей сложного словообразования на узбекском и каракалпакском языках, было внедрено в бюро переводов ООО «Респект» для устранения недостатков в переводе текстов (Министерство информационных технологий и коммуникаций от 19 февраля 2021 года №33-8/1255 - номер справки). В результате разработанное на основе предложенных алгоритмов программное обеспечение позволило переводческой компании «Респект» увеличить полезный рабочий коэффициент общей деловой активности на 90 процентов.

программа, разработанная на основе алгоритма и математической модели составления простых предложений на узбекском и каракалпакском языках, была внедрена в ООО «FARMA PRINT NUKUS» в связи с недостатками, возникающими при переводе простых предложений на узбекский и каракалпакский языки (справочник Министерства по развитию информационных технологий и коммуникаций от 19 февраля 2021 года №33-8/1255). Разработанное на основе предложенных алгоритмов программное обеспечение позволило ООО «FARMA PRINT NUKUS» в результате сократить общий объем работ со временем в 1-2 раза.

программа, разработанная на основе алгоритма и математической модели сложного словообразования в узбекском языке, была внедрена в ООО «ROBOTICS LAB FUTURE TECHNOLOGY COMPANY» с целью создания сложных слов в узбекском языке и устранения недостатков в их переводе (справочник Министерства развития информационных технологий и коммуникаций от 19 февраля 2021 года №33-8/1255). В результате разработанное на основе предложенных алгоритмов программное обеспечение позволило «ROBOTICS LAB FUTURE TECHNOLOGY COMPANY» увеличить скорость обработки текстов на 50% в общей деловой активности.

Апробация результатов исследования. Всего результаты данного исследования обсуждались на 16 конференциях из них 7 международных и 9 республиканских научно-практических.

Публикация результатов исследования. Основные результаты исследования опубликованы в 4 научных работах, 3 из них опубликованы в журналах, рекомендованных Высшей аттестационной комиссией Республики Узбекистан для публикации основных научных результатов докторских

диссертаций, в том числе 1 в зарубежных журналах и 2 в республиканских журналах, 1 в зарубежных журналах а также получены 2 свидетельства об официальной регистрации программ.

Структура и объем диссертации. Диссертация содержит 128 страниц и состоит введения, четырех глав, заключения, списка использованной литературы и приложений.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обосновывается актуальность и необходимость темы диссертации, показано соответствие с приоритетными направлениями развития науки и техники Республики Узбекистан, определяются цели и задачи исследования, определяются объект и предмет исследования, приводится достоверность полученных результатов, информация о состоянии исследования, опубликованных работах и структуре диссертации.

В первой главе диссертации «**Этапы развития компьютерной лингвистике и электронных переводчиков**» представлено современное состояние и перспективы развития электронного переводчика. Во втором параграфе анализируются существующие электронные переводчики, их достоинства и недостатки. В третьем параграфе рассматриваются проблемы создания электронных переводчиков турецкого языка.

При освещении электронного перевода механизма с одного тюркского языка на другой учитывается следующее:

1. В основном в тюркских языках используются кириллица и латиница. В каждом тюркском языке также есть буквы, обозначающие определенные звуки. Например, кроме букв кириллицы в узбекском языке есть еще четыре буквы - буквы ў, қ, ғ, ҳ. В латинском алфавите эти буквы отмечены разными фонетическими символами. Подобные буквы есть и в других тюркских языках. Следует отметить, что даже в одном тюркском языке может использоваться несколько алфавитов. Например, узбеки в Узбекистане используют латиницу, а узбеки в Афганистане – арабский алфавит. Поэтому необходимо будет применить все алфавиты тюркских языков между собой.

2. Для каждого тюркского языка необходимо организовать префиксы и постсуффиксы, то есть префиксы и суффиксы, а также базы данных по словообразовательным базам. Аналогичные базы создаются и для падежей. Эти базы должны быть организованы отдельно на кириллице и латинице.

3. Создание формальной модели языков для каждого тюркского языка и языков в группе.

4. Создание электронного словаря тюркских языков.

5. Создание базы данных для повторного фразирования синонимических единиц в тюркских языках.

6. Создание функционального дополнения программного обеспечения для электронного перевода.

7. Создание модели управления электронным переводом.

8. Тестирование программного обеспечения, верификация и валидация.

Вторая глава диссертации «Математическая модель образования сложных слов для тюркских языков» посвящена разработке математической модели образования сложных слов в узбекском и каракалпакском языках, а построение сложных слов создается путем добавления аффиксов к ядру по следующей схеме.

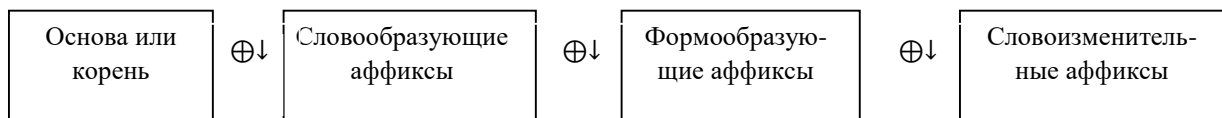


Рис 1. Сложная словообразовательная схема

На основе схемы, приведенной на рис. 1, была разработана математическая модель составления сложных слов в узбекском языке.

$$C = \sum_{i=1}^n x_i \oplus \downarrow \sum_{j=1}^m a_j \oplus \downarrow \sum_{k=1}^l f_k \oplus \downarrow \sum_{r=1}^q p_r$$

Здесь C – сложное слово, $\sum_{i=1}^n x_i$ – основа слова, где каждое слово считается набором x , $\sum_{j=1}^m a_j$ – словообразующие аффиксы, $\sum_{k=1}^l f_k$ – формообразующие аффиксы, $\sum_{r=1}^q p_r$ – словоизменяющие аффиксы, где i, j, k и r – длина слов и аффиксов. Здесь \oplus – операция соединения, \downarrow символ – операция соединение или не соединение.

Приведем пример, основанный на приведенной выше схеме и математической модели создания сложных слов (рис. 2).

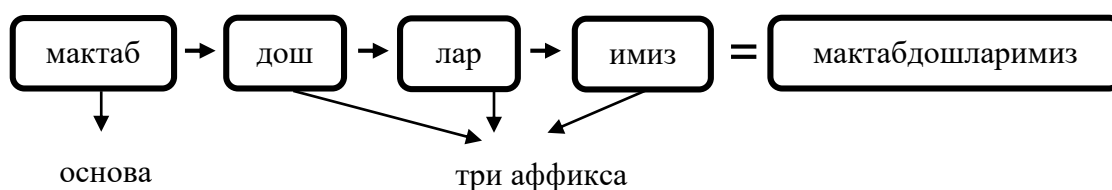


Рис. 2. Пример использования словообразовательных схем в узбекском языке

1. «мактаб» - основа слов, корень
2. «дош» - словообразующий аффикс,
3. «лар» - формообразующий аффикс,
4. «имиз» - словоизменяющий аффикс.

Как видно из примера, слово образуется путем сложения трех слов-суффиксов узбекского языка к корню слова.

Исходя из приведенной выше математической модели, рассмотрим последовательность создания сложных слов в узбекском языке.

$$C_1 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_2 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_3 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_4 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_5 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_6 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{r=1}^q p_r(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

$$C_7 = \sum_{i=1}^n x_i \oplus \begin{cases} \sum_{j=1}^m a_j(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \\ \sum_{k=1}^l f_k(y_i) \sim I(z), & \text{где } z = y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1} \end{cases}$$

Здесь $C_1, C_2, C_3, C_4, C_5, C_6$ и C_7 – основа слов, $I(z)$ – одна местный предикат, каждый суффикс z представлен в виде комбинации двух или более слов $y_0 \wedge y_1 \wedge y_2 \wedge \dots \wedge y_{i-1}$, где i – длина аффикса.

Каракалпакский язык также относится к группе агглютинативных языков, образование сложных слов осуществляется подобно узбекскому языку (рис. 3).

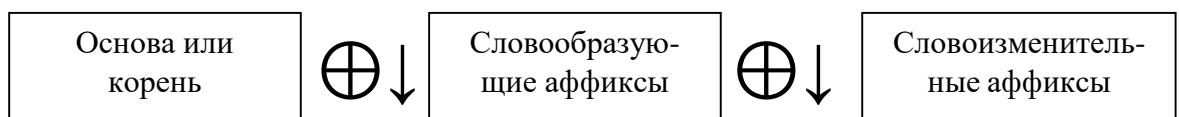


Рис. 3. Схема образования сложных слов в каракалпакском языке

В третьей главе диссертации, называемой «**Последовательность применения алгоритмов семантической модели при построении предложений для двуязычных тюркских языков**» разработаны схема, математическая модель и алгоритм составления простых и сложных предложений в узбекском и каракалпакском языках.

Известно, что в узбекском языке существует обычный порядок слов предложении. В соответствии с этим порядком притяжательное число ставится перед притяжательным, а причастие - после него, определитель предшествует определенному, падежу и дополнительному причастию, причастие обычно стоит в конце предложения.

Однако, чтобы выразить определенный методологический смысл, обычный порядок слов в предложении может быть изменен, такой измененный порядок слов также называется инверсией.

В научном и формальном стилях в основном сохраняется обычный порядок слов в предложении. Но в разговорном, художественном и публицистическом стилях может измениться порядок, то есть инверсия, может применяться.

Простые и сложные предложения в узбекском языке выполняются в следующей последовательности (рис. 4 и 5):

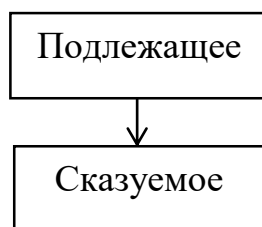


Рис. 4. Схема простых предложений на узбекском языке

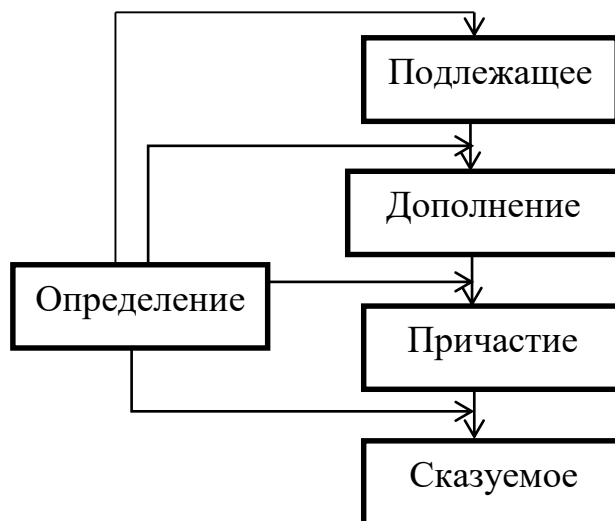


Рис. 5. Схема составления сложных предложений на узбекском языке

На основе приведенных выше схем были разработаны математические модели для образования простых и сложных предложений в узбекском языке.

$$G_{\text{сод}} = A(x_i, \dots, x_n) \cup B(x_i, \dots, x_n)$$

$$G_{\text{мур}} = A(x_i, \dots, x_n) \cup D(x_i, \dots, x_n) \cup S(x_i, \dots, x_n) \cup K(x_i, \dots, x_n) \cup B(x_i, \dots, x_n)$$

Здесь $G_{\text{сод}}$ – простое предложение, $G_{\text{мур}}$ – сложное предложение, $A(x_i, \dots, x_n)$ – подлежащее, $D(x_i, \dots, x_n)$ – дополнение, $S(x_i, \dots, x_n)$ – причастие, $K(x_i, \dots, x_n)$ – определение, $B(x_i, \dots, x_n)$ – сказуемое.

$$G_1 = \left\{ \begin{array}{l} A(x_i, \dots, x_n) \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{array} \right.$$

$$G_2 = \begin{cases} A(x_i, \dots, x_n) \\ K(x_i, \dots, x_n) \rightarrow a_{j(\text{от})} \oplus P_{r(\text{си}\phi)} \oplus P_{r(\phi)} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

$$G_3 = \begin{cases} A(x_i, \dots, x_n) \\ S(x_i, \dots, x_n) \rightarrow P_{r(\text{от})} \oplus P_{r(\text{прав})} \oplus P_{r(\phi)} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

$$G_4 = \begin{cases} A(x_i, \dots, x_n) \\ K(x_i, \dots, x_n) \rightarrow a_{j(\text{от})} \oplus P_{r(\text{си}\phi)} \oplus P_{r(\phi)} \\ D(x_i, \dots, x_n) \rightarrow f_{k(\text{прав})} \oplus P_{r(\text{от})} \oplus P_{r(\phi)} \oplus P_{r(\text{прав})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

$$G_5 = \begin{cases} A(x_i, \dots, x_n) \\ D(x_i, \dots, x_n) \rightarrow f_{k(\text{прав})} \oplus P_{r(\text{от})} \oplus P_{r(\phi)} \oplus P_{r(\text{прав})} \\ S(x_i, \dots, x_n) \rightarrow P_{r(\text{от})} \oplus P_{r(\text{прав})} \oplus P_{r(\phi)} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

$$G_6 = \begin{cases} K(x_i, \dots, x_n) \rightarrow a_{j(\text{от})} \oplus P_{r(\text{си}\phi)} \oplus P_{r(\phi)} \\ A(x_i, \dots, x_n) \\ D(x_i, \dots, x_n) \rightarrow f_{k(\text{прав})} \oplus P_{r(\text{от})} \oplus P_{r(\phi)} \oplus P_{r(\text{прав})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

$$G_7 = \begin{cases} D(x_i, \dots, x_n) \rightarrow f_{k(\text{прав})} \oplus P_{r(\text{от})} \oplus P_{r(\phi)} \oplus P_{r(\text{прав})} \\ B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)} \end{cases}$$

$$G_8 = \{B(x_i, \dots, x_n) \rightarrow f_{k(\phi)} \oplus P_{r(\text{от})} \oplus f_{k(\phi)} \oplus a_{j(\text{от})} \oplus P_{r(\phi)} \oplus f_{k(\phi)}\}$$

В каракалпакском языке простые и сложные предложения составляются так же, как и в узбекском языке (рис. 6 и 7).

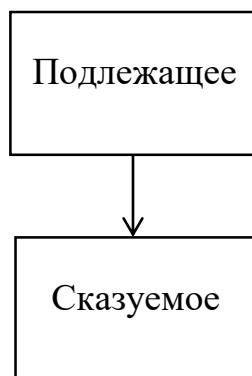


Рис. 6. Схема составления простых предложений на каракалпакском языке

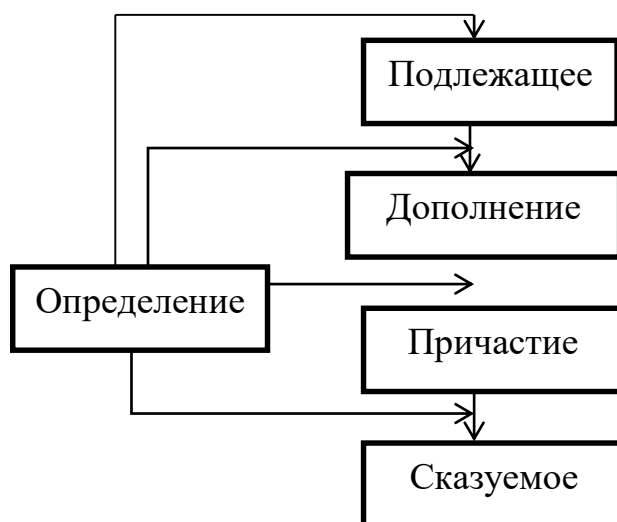


Рис. 7. Схема составления сложных предложений на каракалпакском языке

Четвертая глава диссертации, называемой «Создание электронного словаря и программного комплекса переводчика для двуязычных тюркских языков» разработаны инфологическая модель базы данных электронного переводчика, электронный словарь и функциональный модуль переводчика, а также инструкция по использованию программы.

Между объектами рассматриваемой предметной области существуют следующие типы связей:

1. 1:1 – связь между объектами и атрибутами.

Таким образом, в рассматриваемом случае в основном идентифицируются объекты и обнаруживается связь между ними, классифицируется (рис. 8).

1. Вводится текст для перевода.
2. Лексикографический анализ текста на языке перевода:
Определить лингвистические и семантические атрибуты.
3. Морфосемический анализ слова в переводимом языке:
Определить семантические отношения и анализ текста.
4. Семантико-синтаксический анализ предложений в переводимом языке.
5. Построить онтологию предложений с учетом семантических атрибутов и привязок.
6. Поиск эквивалентов семантических структур в переводимом языке.
7. Формирование семантико-синтаксической структуры предложения в переводном языке.
8. Формирование словообразовательных и словоизменительных средств в переводном языке: синтез многозначных слов на основе контекста.
9. Переведенный текст.

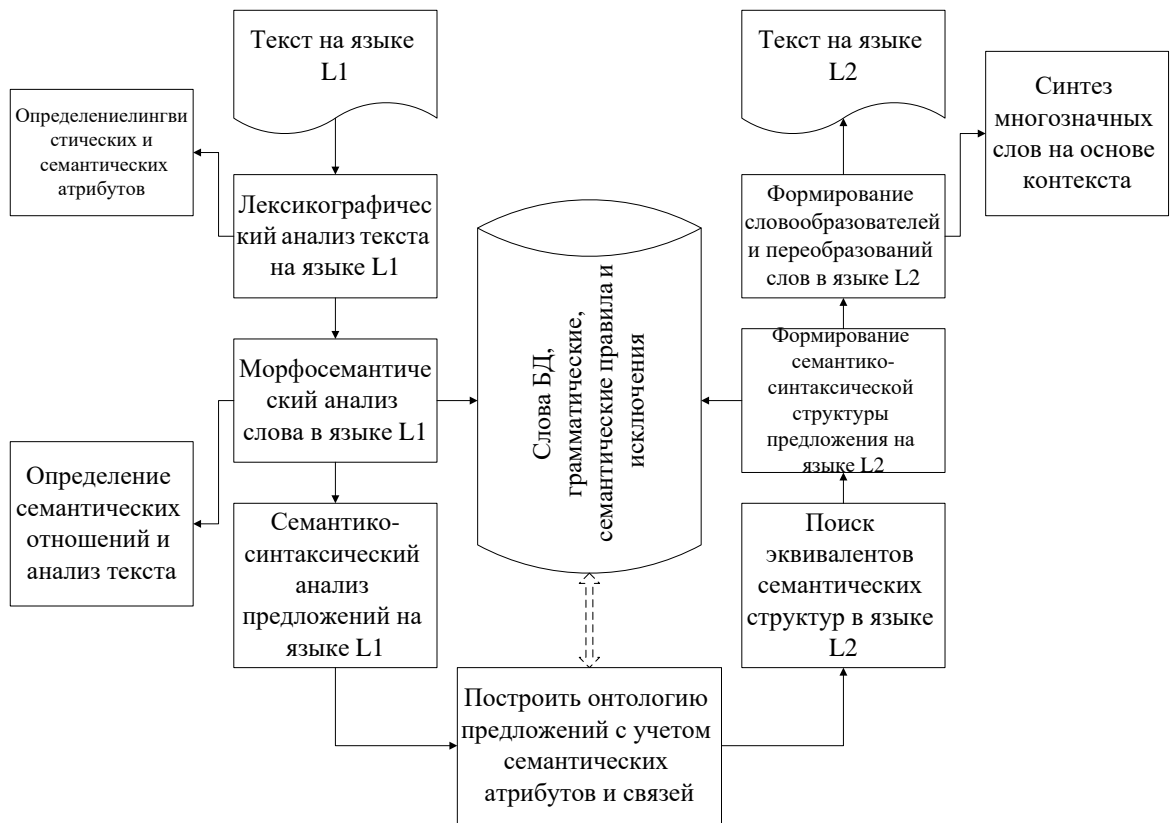


Рис. 8. Семантический анализ и схема машинного перевода генераторной системы с узбекского на каракалпакский

Схема электронного преобразователя IDEF0 представлена на рис. 9.

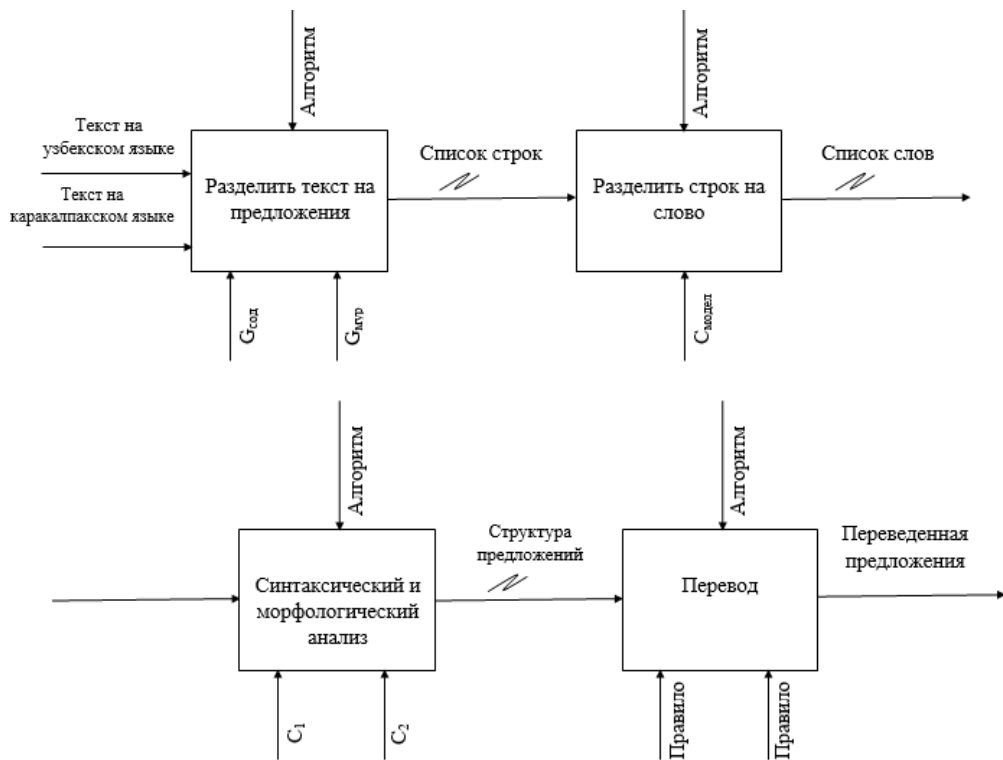


Рис. 9. Схема электронного переводчика IDEF0

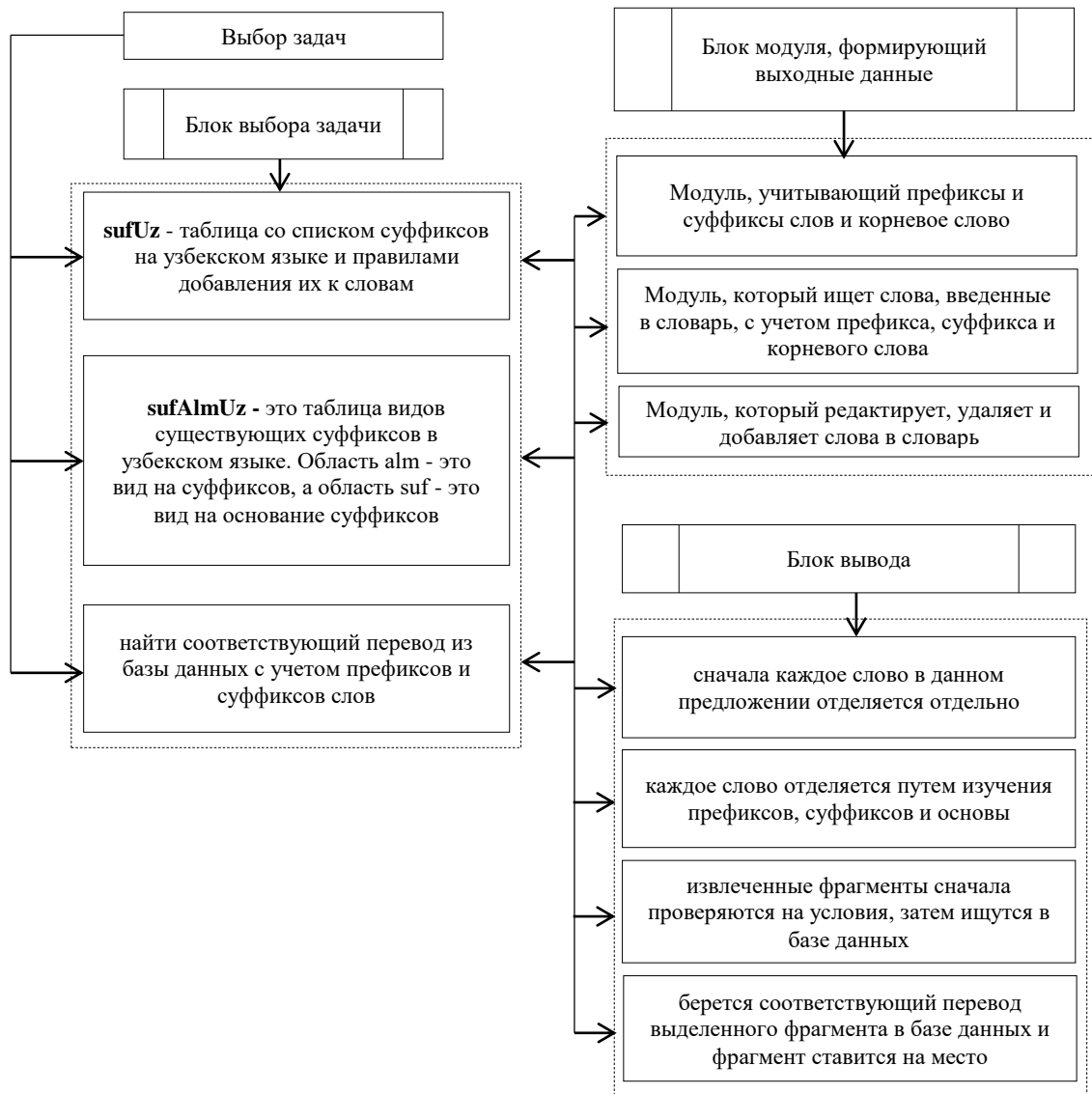


Рис. 10. Функциональная схема работы ПО

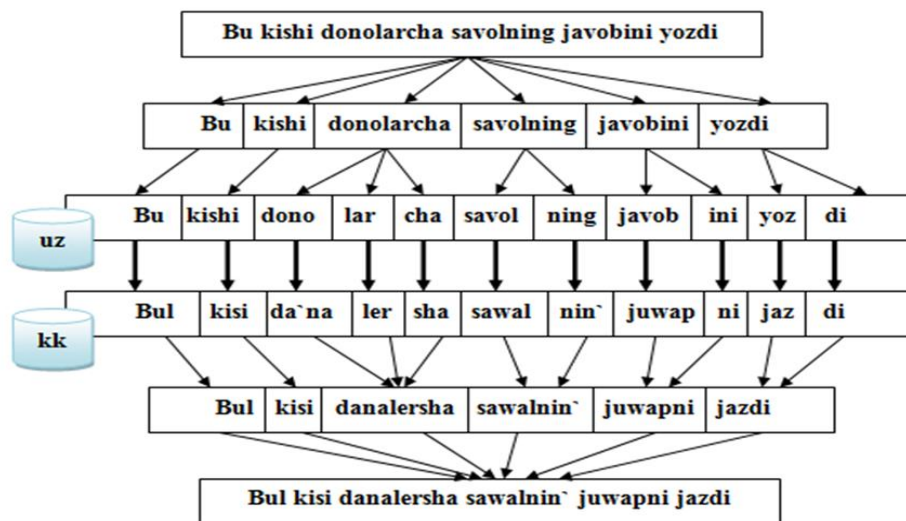


Рис. 11. Алгоритм перевода на основе семантического анализа

Извлеченные префикс, основа и суффикс проверяются из базы данных соответственно. После проверки новое слово будет выделено и опубликовано. Если слово не найдено из таблицы, то оно печатается как текст перевода (рис. 10-12).

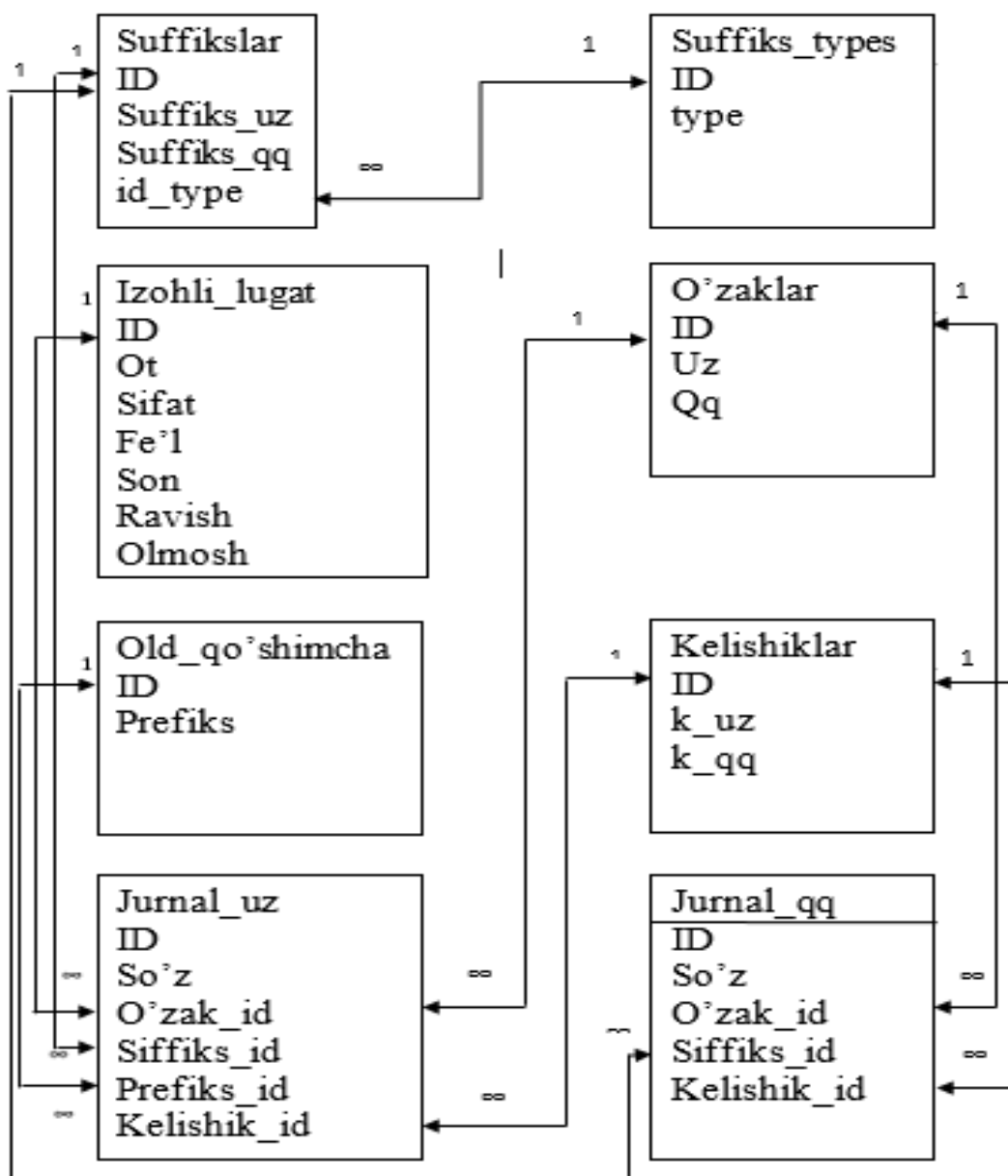


Рис. 12. IDEF1x схема электронного переводчика

Если слова, используемые в переводе, не найдены в базе данных, она предупредит пользователя или выдаст то же самое слово.

ЗАКЛЮЧЕНИЕ

В результате исследовательской работы над диссертацией на тему «Модели и алгоритмы электронного перевода для тюркских языков (на примере перевода с узбекского языка на каракалпакский язык» представлены следующие выводы:

1. Разработана математическая модель лексико-морфологического анализа и построения сложной лексики узбекского и каракалпакского языков, позволяющая формировать простые и сложные слова на узбекском и каракалпакском языках, на основе которой разработана электронная словарная база и база знаний по узбекским и каракалпакским языкам;
2. На основе математической модели сложного словообразования для узбекского и каракалпакского языков был создан эффективный алгоритм, позволивший вывести последовательность словообразования в узбекском и каракалпакском языках, на основе которого было разработано программное средство для создания электронного переводчика для тюркских языков;
3. На основе двуязычного электронного словаря и базы данных переводчиков разработаны алгоритмы поиска слов и определения взаимного перевода текстов, проведен процесс перевода с узбекского языка на каракалпакский с использованием семантического анализа, на основе этого семантического анализа построено дерево решений слов;
4. Решена задача разработки IDEF0-модели программного комплекса электронного переводчика тюркских языков, функциональной структуры программных модулей и инфологической IDEF1x-модели базы данных и объектно-ориентированных моделей базы знаний. Определены модули программного комплекса электронного переводчика для тюркских языков и его компоненты, структура базы данных, структура базы знаний и выполняемая в ней работа;
5. На основе предложенных моделей и алгоритмов был создан программный интерфейс и удобный пользовательский интерфейс, который позволит переводить двуязычные узбекско-каракалпакские и каракалпакско-узбекские тексты. Программное обеспечение служит для быстрой и качественной реализации перевода для пользователей;
6. Разработанное программное обеспечение позволило в 1-2 раза сократить общий объем деятельности бюро переводов «Респект», ООО «FARMA PRINT NUKUS» и ООО «ROBOTICS LAB FUTURE TECHNOLOGY COMPANY».

**SCIENTIFIC COUNCIL AWARDING SCIENTIFIC DEGREES
DSc.13/30.12.2019.T.07.01 AT TASHKENT UNIVERSITY OF
INFORMATION TECHNOLOGIES**

TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES

ABIDOVA SHAKHNOZA BAHODIROVNA

**MODELS AND ALGORITHMS FOR ELECTRONIC TRANSLATION FOR
TURKIC LANGUAGES (ON THE EXAMPLE OF TRANSLATION FROM
UZBEK INTO KARAKALPAK)**

05.01.04 – Mathematical and software support of computers, complexes and
computer networks

**DISSERTATION ABSTRACT OF THE DOCTOR OF PHILOSOPHY
(PhD) ON TECHNICAL SCIENCES**

Tashkent – 2021

The theme of dissertation of doctor of philosophy (PhD) on technical sciences was registered at the Supreme Attestation Commission at the Cabinet of Ministers of the Republic of Uzbekistan under number B2021.1.PhD/T2080.

The dissertation has been prepared at Tashkent university of information technologies named after Mukhammad al-Kharizmi.

The abstract of the dissertation is posted in three languages (Uzbek, Russian, English (resume)) on the website (www.tuit.uz) and on the website of «Ziyonet» Information and Educational portal (www.ziyonet.uz).

Scientific adviser:

Nuraliev Fakhridin

Doctor of Technical Sciences, Docent

Official opponents:

Mukhamedieva Dilnoz Tulkunovna

Doctor of Technical Sciences, Professor

Norov Abdisait Muradovich

Doctor of Philosophy of Technical Sciences (PhD)

Leading organization: Tashkent State Technical University named after Islam Karimov

The defense will take place «19» june 2021 at 9:00 on the meeting of Scientific Council No DSc.13/30.12.2019.T.07.01 at Tashkent University of Information Technologies named after Mekhammad al-Kharizmi (Address: 100202, Tashkent city, Amir Temur Street, 108. Tel.: (99871) 238-64-43, fax: (99871) 238-65-52, e-mail: tuit@tuit.uz).

The dissertation can be reviewed at the Information Resource Centre of Tashkent University of Information Technologies named after Mekhammad al-Kharizmi (is registered under No 209). (Address: 100202, Tashkent city, Amir Temur Street, 108. Tel.: (99871) 238-65-44; fax: (99871) 238-65-52).

Abstract of dissertation sent out on «15» june 2021 y.
(mailing report No 17 on «10» june 2021 y.).



R.Kh. Khamdamov
Chairman of the scientific council
awarding scientific degrees,
Doctor of Technical Sciences, Professor

E.Sh. Nazirova
Scientific secretary of scientific council
awarding scientific degrees,
Doctor of Technical Sciences, Docent

M.A. Rakhmatullaev
Chairman of the academic seminar under the
scientific council awarding scientific degrees,
Doctor of Technical Sciences, Professor

INTRODUCTION (abstract to PhD dissertation)

The aim of the research work construction of a model and algorithms for translation from Uzbek into Karakalpak, a bilingual electronic dictionary database and development of a software package.

The objects of the research work electronic translators who translate from Uzbek into Karakalpak make up the linguistic corpus of the Uzbek and Karakalpak languages.

The scientific novelty of the research work:

on the basis of the linguistic models of the Uzbek-Karakalpak languages, an electronic dictionary base of the Turkic languages was created;

on the basis of the lexical and morphological analysis of the Uzbek and Karakalpak languages, a mathematical model and an algorithm for the formation of complex vocabulary and a simple sentence structure have been developed;

a bilingual electronic dictionary of the Uzbek and Karakalpak languages was developed, as well as an interpreter algorithm for finding words and determining the mutual translation of texts based on data and a knowledge base;

functional modules of programming processes were developed on the basis of modular analysis of algorithms for electronic translation into Uzbek and Karakalpak languages;

on the basis of functional modules of programming processes for electronic translation algorithms into Uzbek and Karakalpak languages, the structure of the software package was developed and a user interface was created.

Implementation of the research results: Based on algorithms and software developed as a result of research conducted within the framework of the dissertation:

the software developed on the basis of algorithms and mathematical models of complex word formation in the Uzbek and Karakalpak languages was implemented in the translation agency "Respect" LLC to eliminate shortcomings in the translation of texts (Ministry of Information Technologies and Communications No. 33/1255 of February 19, 2021). As a result, the software developed on the basis of the proposed algorithms allowed the translation company "Respect" to increase the useful working coefficient of total business activity by 90 percent.

The program, developed on the basis of an algorithm and a mathematical model for composing simple sentences in Uzbek and Karakalpak languages, was implemented in FARMA PRINT NUKUS LLC due to shortcomings that arise when translating simple sentences into Uzbek and Karakalpak languages (reference book of the Ministry for the Development of Information Technologies and Communications No. 33/1255 of February 19, 2021). As a result, the software developed on the basis of the proposed algorithms allowed FARMA PRINT NUKUS LLC to reduce the volume of total business activity by 1-2 times compared to the time.

The program, developed on the basis of an algorithm and a mathematical model of complex word formation in the Uzbek language, was implemented in ROBOTICS LAB FUTURE TECHNOLOGY COMPANY LLC in order to create complex words in the Uzbek language and eliminate shortcomings in their

translation (reference book of the Ministry of Information Technology and Communications Development No. 33/1255 of February 19, 2021). As a result, the software developed on the basis of the proposed algorithms allowed "ROBOTICS LAB FUTURE TECHNOLOGY COMPANY" to increase the speed of text processing by 50% in the overall business activity.

Structure and volume of the dissertation: The dissertation contains 128 pages and consists of an introduction, four chapters, a conclusion, a list of references and appendices.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I бўлим (I часть; I part)

1. Абидова Ш.Б. Электрон таржимон яратишда ўзбек тилининг морфологик таҳлилини амалга оширишнинг инфологик модели. // Мухаммад ал-Хоразмий авлодлари, № 3 (13), сентябр 2020. –Б. 131-136. (05.00.00; №10).
2. Нуралиев Ф.М., Абидова Ш.Б. Икки тилнинг семантик таҳлили. ТАТУ хабарлари журнали. Т.: 2017, №3(40). –Б. 113-117. (05.00.00; №31).
3. Nazirova E. Sh., Abidova Sh.B. MATHEMATICAL MODEL AND ALGORITHM FOR CALCULATING COMPLEX WORDS IN THE KARAKALPAK LANGUAGE. “Bulletin of TUIT: Management and Communication Technologies”. 2019, 1 (44). –P. 1-7. (30.07.2020; №283/7.1).
4. Nazirova E.Sh., Abidova Sh.B. Mathematical model and algorithm for calculating complex words int Uzbek Language. // Solid State Technology. Volume: 64, Issue 2. 2021 y. pp. 4347-4359.

II бўлим (II часть; II part)

5. Абидова Ш.Б., Абдукаримов С. Электронные переводчики. // “Илм ва фан равнақида олима аёллар ўрни” мавзусида Ўзбекистон Республикаси Олий ва ўрта махсус таълим вазирлиги миқёсида илмий - амалий анжуман. 20-21 ноябр 2014 йил. –Б. 166-167.
6. Абидова Ш.Б., Махмудов А.З. Типы электронных переводчиков. // “Илм ва фан равнақида олима аёллар ўрни” мавзусида Ўзбекистон Республикаси Олий ва ўрта махсус таълим вазирлиги миқёсида илмий - амалий анжуман. 20-21 ноябр 2014 йил. –Б. 243-245.
7. Абидова Ш.Б., Махмудов А.З. Создания программного обеспечения электронного переводчика для тюркских языков. // XVМеждународной конференции «Информатика: проблемы, методология, технологии»г. Воронеж, 12-13 февраля 2015 г. 229-232 стр.
8. Абидова Ш.Б. Туркий тиллар учун электрон луғат яратиш муаммолари. // Иқтисодиётнинг реал тармоқларини инновацион ривожланишида ахборот-коммуникация технологияларининг аҳамияти. Республика илмий-техник анжумани. Тошкент 2017 й. 6-7 апрел. –Б. 267-269.
9. Абидова Ш.Б. Особенности морфологии узбекского языка создании электронного узбекско-каракалпакского словаря-переводчика. // «Актуальные проблемы математики и механики – SAWMA-2018» тезисы докладов республиканской научно-практической конференции с участием зарубежных женщин-ученых. Хива, 2018 г., 25-26 октябр. 4-5 стр.
10. Абидова Ш.Б. “Лингвистическая схема трансферной системы двуязычного перевода”. // Тезисы международной конференции “Актуальные

проблемы прикладной математики и информационных технологий”. Тошкент, 2019 г., 14-15 ноябр. 283-284 стр.

11. Абидова Ш.Б. Таълим сифатини оширишда электрон таржимонларнинг ўрни. // Юқори малакали мутахассис-кадрларни тайёрлашда ўқитишнинг кредит тизимини қўллаш масалалари. Республика илмий-услубий конференцияси маърузалар тўплами. Тошкент 2020 й. –Б. 133-134.

12. Нуралиев Ф.М., Абидова Ш.Б. Роль морфологии узбекского языка для создания электронного переводчика. // Математик моделлаштириш, алгоритмлаш ва дастурлашнинг долзарб муаммолари” республика илмий-амалий конференцияси материаллар тўплами. Тошкент, 2018 й., 17-18 сентябр. –Б. 373-375.

13. Нуралиев Ф.М., Абидова Ш.Б. Роль морфологии при переводе с одного языка на другой. // Материалы докладов 84-й научно-технической конференции, посвященной 90-летию юбилею БГТУ и Дню белорусской науки (с международным участием). Минск, 3-14 феврал 2020 г. 366-369 стр.

14. Нуралиев Ф.М., Абидова Ш.Б., Холиқулова Х.С. ЭЛЕКТРОН ТАРЖИМОН ЯРАТИШДА СЎЗЛАРНИНГ СЕМАНТИК МОДЕЛИ. // “Инновацион ғоялар, ишланмалар амалиётга: муаммолар ва ечимлар” Халқаро илмий-амалий онлайн анжуман. Андижон 2020 й. 27-28 май. –Б. 199-200.

15. Рахмонов Қ.С., Абидова Ш.Б. Электрон таржимон яратишдаги муаммолар. // “Ахборот технологиялари ва телекоммуникация тизимларини самарали ривожлантириш истиқболлари”. Республика илмий-техник анжумани. 12-13 март. 2015. – Б. 253-257.

16. Abidova Sh.B., Abdulkarimov S.S. Strategy of machine translation. // Transactions of the international scientific conference “Perspectives for the development of information technologies ITPA 2015” 4-5 november, 2015 year. –P. 248-250.

17. Nuraliyev F.M., Abidova Sh.B. Conceptual model of the uzbek-karakalpak machine translation system. // Fundamental matematika muammolari va ularning tatbiqlari, Navoiy 25 may 2019 y. –B. 4-5.

18. Нуралиев Ф.М., Абидова Ш.Б. Икки тилли электрон луғат маълумотлар базаси. // Ўзбекистон Республикаси интеллектуал мулк агентлигидан гувоҳнома ВГУ00391, 18.01.2020 й.

19. Нуралиев Ф.М., Назирова Э.Ш., Абидова Ш.Б., Хожиев С.А. “Икки тилли электрон луғат (ўзбек-қорақалпоқ)”. // Ўзбекистон Республикаси интеллектуал мулк агентлигидан гувоҳнома DGU07829, 29.02.2020 й.

20. Nazirova E.Sh., Abidova Sh.B., Hojiev S.A., Jamolov N.F. Construction of a mathematical model of complex word formation in the Uzbek-Karakalpak language. // Scientific horizon in the context of social crises. Proceedings of the 7th International Scientific and Practical Conference. Tokyo-Japan, 6-8.02.2021 y. pp.1117-1123.

Автореферат «ГАТУ хабарлари» илмий-техника ва ахборот-таҳлилий
журнали таҳририятида таҳрирдан ўтказилди ҳамда ўзбек, рус ва инглиз
тилларидаги матнларининг мослиги текширилди.