

**К.КАРАСАЕВ АТЫНДАГЫ БИШКЕК ГУМАНИТАРДЫК
УНИВЕРСИТЕТИ
И.АРАБАЕВ АТЫНДАГЫ КЫРГЫЗ МАМЛЕКЕТТИК
УНИВЕРСИТЕТИ
КЫРГЫЗ РЕСПУБЛИКАСЫНЫН УЛУТТУК ИЛИМДЕР
АКАДЕМИЯСЫНЫН
Ч.АЙТМАТОВ АТЫНДАГЫ ТИЛ ЖАНА АДАБИЯТ ИНСТИТУТУ**

Д 10.17.548 Диссертациялык кеңеш

Кол жазма укугунда

УДК: 811.512.154(575.2)(04)

САРЫГУЛ КЕРИМ

**КЫРГЫЗ ТИЛИНИН ЛЕКСИКОГРАФИЯЛЫК МААЛЫМАТ
БАЗАСЫН ТҮЗҮҮ МАСЕЛЕСИ**

Адистиги 10.02.01 – кыргыз тили

Филология илимдеринин кандидаты окумуштуулук даражасын
изденип алуу үчүн жазылган диссертациянын

А В Т О Р Е Ф Е Р А Т Ы

Бишкек – 2018

Диссертациялык иш, Кыргыз Республикасынын Улуттук илимдер академиясынын Ч.Айтматов атындагы Тил жана адабият институтунда аткарылды.

Илимий жетекчи:

филология илимдеринин доктору, профессор
Токоев Таалайбек Токтосунович

Расмий оппоненттер: филология илимдеринин доктору, профессор
Садыков Талполот Садыкович

филология илимдеринин кандидаты, доцент
Абдыразакова Гульмира Шарапидинова

Жетектөөчү мекеме: К.Тыныстанов атындагы Ысык-Көл мамлекеттик университетинин кыргыз тил илими кафедрасы.
Дареги: 722200, Каракол ш., Ж.Абдырахманов көч., 103.

Диссертациялык иш 2018-жылдын 8-июнунда саат 10:00дө К.Карасаев атындагы Бишкек гуманитардык университети, И.Арабаев атындагы Кыргыз мамлекеттик университети жана Кыргыз Республикасынын Улуттук илимдер академиясынын Ч.Айтматов атындагы Тил жана адабият институтуна караштуу филология илимдеринин доктору (кандидаты) окумуштуулук даражасын коргоо боюнча түзүлгөн Д 10.17.548 Диссертациялык кеңештин жыйынында корголот.

Дареги: 720044, Бишкек ш., Тынчтык проспекти, 27.

Диссертациялык иш менен К.Карасаев атындагы БГУнун илимий китепканасынан таанышууга болот.

Дареги: 720044, Бишкек ш., Тынчтык проспекти, 27.

Автореферат 2018-жылдын 7-майында жөнөтүлдү.

Диссертациялык кеңештин
окумуштуу катчысы, филология
илимдеринин кандидаты,
доцент



Каратаева С.

ИШТИН ЖАЛПЫ МҮНӨЗДӨМӨСҮ

Теманын актуалдуулугу. Жалпы тил илими, лексикография, фонология ж.б. тил илиминин көп түрдүү тармактарында компьютерлерге, т.а., системалардын электрондоштурулушуна жана каалаган натыйжага кыска убакытта жетүүгө муктаждык пайда болду.

Дүйнөдө азыркы күнгө чейин миңдеген компьютердик программалар түзүлүп, бул иш-аракеттер келечекте да ылдам темп менен өнүгөт. Ошондой болсо да, тил илимине арналган программалардын саны аз.

Акыркы жылдарда сөздүктөрдүн жана тексттердин электрондук системаларга өткөрүлүп, колдонууга сунушталышы менен бирге, бул тармакта айрым өзгөрүүлөр жүрсө да, иш-аракеттердин көпчүлүгү мурдатан колдонулуп жүргөн китеп түрүндөгү сөздүктөрдүн, же айрым тексттердин иштелбеген абалында электрондук системага киргизилип иштетилишинин негизинде жүрөт. Мындай иш-аракеттердин лингвистика, лексикография, текстология, фонология, семантика ж.б. тармактар үчүн таасири аз. Ал маалыматтар колдонуучуларга маалыматты издеп табуу учурунда убакытты үнөмдөөгө шарт түзөт.

Маалыматтардын электрондук системаларга өткөрүлүшү эң маанилүү иш-аракеттерден болуп саналат. Ошентсе да ал маалыматтарды толук жана иштеле тургандай абалга алып келүү абзел.

Кыргыз тилинин кең форматтагы маалымат базасын түзүү, анын эрежелерине ылайык диаграммаларды жана алгоритмдерди даярдоо лингвистикалык жактан да, лексикографиялык жактан да тилдин азыркы күндөгү технологияларга жана керектөөлөргө жараша эң мыкты деңгээлде жооп беришинен улам маанилүү.

Диссертациянын темасынын илимий программалар менен болгон байланышы. Диссертациялык иштин темасы Кыргызстандагы жалпы илимий жана мамлекеттик изилдөө программаларына шайкеш келип, жекече аткарылган илимий эмгек болуп эсептелет.

Изилдөөнүн максаты жана милдеттери. Изилдөөнүн максаты кыргыз тилинин лексикографиялык маалымат базасын түзүү.

Бул максатка жетүү үчүн төмөнкүдөй **милдеттер** коюлду:

1. Кыргыз тилинин сөз байлыгынын кең форматтуу маалымат базасынын, кыргыз тилинин эрежелерине ылайык түзүлгөн диаграммалар жана алгоритмдердин даярдалышы;

2. Аталган диаграмма жана алгоритмдердин даярдалышы тил илими (*айрыкча компьютердик лингвистика*), лексикография (*айрыкча электрондук лексикография*) үчүн канчалык деңгээлде маанилүү экенин аныктоо;

3. Бул типтеги иш-аракеттер жасалганда, маалыматтардын компьютер технологиялары жана программалоо тилдери тарабынан кандайча иштей турганын жана кайда колдонула турганын баяндоо;

4. Жасала турган иш-аракеттердин натыйжасында пайда болгон маалыматтар тил илиминин бардык тармактарында колдонулуу менен катар, башка илимдерге да булак катары кызмат кыларын аныктоо.

Изилдөөнүн илимий жаңылыгы. Кандайдыр бир тилдин керектөөлөрүнө жана технологиялык мүмкүнчүлүктөрүнө жооп бериши, түрдүү тармактарда натыйжалуу колдонулушу тилчилер менен колдонуучулардын көз карашынан алганда, аткарыла турган иш-аракеттер үч баскычтан турат:

1. Өркүндөтүлгөн сөздүк маалымат базасын түзүү;
2. Тилдик эрежелердин негизинде алгоритмдерди жана диаграммаларды түзүү;
3. Программаларды түзүү.

Сөздүк маалымат базасы түзүлүп жатканда, сөздөрдүн кайсы өзгөчөлүгү эске алынганы, алардын маалымат базасында кандай сакталганы, тилдик алгоритмдер менен диаграммалардын лингвистикалык мааниси эске алынуу менен, үлгүлөр камтылып, бул тармакта иш алып бара турган изилдөөчүлөргө багыттар көрсөтүлдү.

Иштин практикалык наркы. Иште колдонулган илимий жаңылыктар бүгүнкү күнгө чейин кыргыз тил илиминде аткарыла элек, ошондуктан алардын илимий негизин түзүү илимий жаңылык болуп эсептелет. Жаңы түзүлгөн программаны тилдин кайсы гана тармагында болбосун, бирдей колдонууга болот. Негизинен, үч функцияны (*анализ, оңдоп-түзөө, иштеп чыгуу*) жүзөгө ашыра турган программалардын колдонулуу чөйрөсү төмөнкүчө: лексикография / электрондук лексикография; лингвистика / компьютердик лингвистика; тил үйрөтүү; текст анализи; жөндөлүш/сүйлөм/текст түзүү; котормо иштери; автоматизациялык системалар; робот технологиялары; интернет ж.б.

Коргоого сунушталган негизги жоболор. Изилдөөгө алынган маалыматтардын негизинде программисттер менен тилчилер бирге иш алып барат. Анын натыйжасында программа:

- Кыргыз тили колдонулуп жаткан электрондук системаларды көзөмөлгө алат. Электрондук системалардагы тилдин тууралыгын текшерет жана колдонуу учурунда айрым эскертүүлөр менен колдонуучулардын кыргызчаны туура колдонушуна шарт түзөт.

- Өркүндөтүлгөн сөздүк маалымат базасы аркылуу көптөгөн жаңы программаларды түзүүгө жана ал маалыматтардын тил илиминен тышкары да тил менен катыштыгы болгон бардык тармактарда колдонууга көмөктөшөт.

- Кыргыз тили менен программалоо тилдеринин логикасын колдонуп, тил үйрөнүү жана үйрөтүү процессин тездетет.

- Маалымат базасындагы сөздөрдү колдонуп, тексттин үстүнөн анализ жүргүзөт. Тексттер тиешелүү болгон тармактар аныкталат, сөз жыштыктары такталат жана колдогу маалыматтардын негизинде текст бааланат.

- Туура иштеген алгоритмдердин негизинде компьютердин сүйлөм түзүшүн камсыз кылат.

- Автоматтык системалардын үндүк буйруктарды кабыл алуу мүмкүнчүлүктөрү кеңейиши мүмкүн.

Жогоруда тизмектелген функциялар негизги гана максаттар болуп саналат. Булардан тышкары да жүздөгөн программаларды түзүүгө жана айрым программаларды өркүндөтүүгө жетише алабыз. Түрк тилдеринин жалпы структуралары, жалпысынан, бирдей болгондуктан, бир тил үчүн даярдалган программаларга айрым өзгөртүүлөрдү киргизүү менен башка тилдер үчүн да колдонсо болот.

Изилдөөчүнүн жекече салымы. Диссертациялык иштеги илимий жаңылыктар, тыянактар изденүүчүнүн жеке талдоолорунун, иликтөөлөрүнүн жана изилдөөлөрүнүн жыйынтыгы болуп саналат.

Кыргыз тилинин сөз байлыгы өркүндөтүлгөн маалымат базасында кандайча орун алышы керек экени аныкталды.

Сөз түркүмдөрүнүн схемалары жана бул схемалардын негизинде сөз кошуу формалары түзүлдү.

Жөнөкөй маалымат базасы менен өркүндөтүлгөн маалымат базасынын ортосундагы айырмачылыктар жана колдонуу чөйрөлөрү изилденди.

Кыргыз тилин азыркы күндүн технологияларында кеңири түрдө колдонуу үчүн маалымат базасынын үлгүсү жана алгоритмдер түзүлүп, бул тармакта иштей турган изилдөөчүлөргө жол көрсөтүлдү.

Иште колдонулган методдор: Изилдөөдө математикалык, салыштырма, сыпаттама, сурамжылоо, байкоо жүргүзүү, системалаштыруу, моделдештирүү сыяктуу методдор колдонулду.

Изилдөөнүн апробацияланышы. Диссертациялык изилдөөнүн негизги жыйынтыктары иштин мазмунун чагылдырган 8 макала жарыяланды.

Диссертациянын структурасы жана көлөмү

Илимий эмгек 251 беттен, 14 схемадан, 10 формадан, 47 сүрөттөн жана 1 программадан турат.

ИШТИН НЕГИЗГИ МАЗМУНУ

Киришүүдө ишке жалпы мүнөздөмө берилип, иликтөөгө алынган проблеманын актуалдуулугу, максаты жана милдеттери, илимий жаңылыгы, теориялык-практикалык мааниси, коргоого коюлган негизги жоболор, колдонулган методдор, диссертациянын апробацияланышы, структурасы жана көлөмү белгиленди.

Ишибиздин **биринчи бабы «КОЛДОНМО ЛИНГВИСТИКАНЫН КАЛЫПТАНЫШЫ ЖАНА ӨНҮГҮШҮ»** деп аталып, беш бөлүктөн турат да, анда колдонмо лингвистика жөнүндө жалпы маалымат, колдонмо тил илиминин түркологияда калыптанышы жана азыркы абалы, түрк тил илиминде колдонмо лингвистика багытынын изилдениш абалы, колдонмо тил илиминин кыргыз жана түрк тилинде изилдениши, лексикографиянын тарыхы жана негизги өзгөчөлүктөрү боюнча илимий сереп камтылды.

Тилдин колдонмо тармагы өзүнүн өтө кең жана ар тараптуу өзгөчөлүктөрү менен айырмаланып, анын калыптануу жана өнүгүү процесси XX-XXI кылымдарды өз ичине камтып турат. Анын байыртадан келе жаткан салттуу негизи – жазуу (*графика*), эне тил менен чет тилин окутуу методикасы жана тилдин лексикографиялык жүйөлөрү экенин өзгөчө белгилеп кетүү зарыл. Тилдин өнүгүш баскычтарына назар салсак, котормо таануу, дешифровка (*шифртаануу*), орфография, транслитерация, терминология боюнча эмгектер тилдин практикалык маселелерин чечүүгө багытталып, колдонмо тил илими менен тыгыз байланыштуу чечиле баштаганын баамдоого болот.

Тилди так илимдер менен байланышта изилдөө маселесине алгачкылардан болуп кайрылган Бодуэн де Куртенэ тил илиминде сандык, математикалык амалдарды кеңири колдонуп, аны мүмкүн болушунча, так илимдерге жакындатуу керек экенин өткөн кылымда эле белгилеп кеткен [Бодуэн де Куртенэ, 1963].

Учурда ар кайсы илим тармагында илимий жана техникалык терминдерди колдонуу кажеттиги кеңири кулач жаюуда. Мындай маселе лексикография тармагынан тышкары өз чечилишин таппагандыгы да маалым. Ошондуктан коомдо компьютердин өтө кеңири колдонулушу менен байланыштуу автоматтык лексикография тармагы өз ордун татыктуу ээлеп келүүдө.

Колдонмо лингвистика – философия, психология, физиология, математика, физика, логика, информатика илимдери менен тыгыз байланыштуу тил илиминин өтө татаал тармагы. Көп тармактуу тил илиминин бул тармагын окумуштуулар «компьютердик лингвистика», «инженердик лингвистика», «автоматтык лингвистика», «эсептөөчү лингвистика», «квантативдик лингвистика», «статистикалык лингвистика» терминдери менен атап келишти.

Колдонмо лингвистиканын башаттары XIX кылымдын башында эле колго алынса да, анын илимий-теориялык маселелери XX кылымдын экинчи

жарымынан баштап түптөлүп, тил илиминин өз алдынча келечектүү тармагы катары бөлүнүп чыгышына өбөлгө түздү.

Тилдин бул тармагынын изилдөө чөйрөсү бүгүнкү күндө өтө кеңири. Анын алгачкы саамалыгы 1957-жылдан тарта «Тил статистикасы» багытындагы изилдөө иштеринен башталып, анын башатында америкалык тилчи Г.Глисон, орус тилчиси Р.Г.Пиотровскийлер дескриптивдик жана статистикалык тил илими тармактарына негиз салышты [Глисон,1959; Пиотровский, 1959].

Түрк тилдеринде колдонмо лингвистика багытына өткөн кылымдын 60-жылдарынан тартып өзгөчө көңүл бурулуп, жалпы Союздук «Тил статистикасы» тармагы боюнча Р.Г.Пиотровский, П.М.Алексеевдер башында турган Ленинград тил мектебинин окуучулары тарабынан жүзөгө ашырыла баштады. 1968-жылы «Абай тили сөздүгүнүн» түзүлүшүнөн тартып, статистикалык тил илими Казакстанда интенсивдүү өнүктү.

Казак тилинин колдонмо тил илиминин өнүгүшү К.Б.Бектаев, Д.А.Байтанаева, А.К.Жубановдун ысымы менен да байланыштуу болсо, өзбек тилинде негизделиши С.А.Ризаевдин, татар тилинде К.Р.Галиуллин, А.А.Тухватуллина, Л.С.Шарипниязова, Р.Г.Бухарев, Б.С.Сафиуллина, Д.Ш.Сулейманов, Азербайжанда К.В.Велиева, Д.Жабарлы, М.Махмудов, А.Фатуллаев ж.б. ысымы менен байланыштуу.

Кыргыз тил илиминде колдонмо лингвистика багыты КР УИАнын Тил жана адабият институтунун Эксперименталдык фонетика лабораториясынын ачылышы менен башталды. Бул мезгилде кыргыз тилинин тыбыштык системаларына тажрыйба жүргүзүлүп, анын жыйынтыктары Т.К.Ахматов, Ж.К.Сыдыков, А.О.Орусбаев, Т.Чымбаев, Т.Садыковдордун эмгектеринде чагылдырылды.

Жаш кыргыз тилчилеринин 70-жылдарда Москва, Ленинград шаарларына аспирантура жана докторантурага жиберилиши тил илиминин бул тармагында тажрыйбалуу адистер менен толукталышына шарт түздү. Тил маселелерин математикалык метод аркылуу изилдөө жана ыктымалдуулук теориясы менен бекемдөөгө байланыштуу алгачкы эмгек профессор С.И.Ибрагимовдун калемине таандык.

Окумуштуу алгачкы жолу математикалык методун кыргыз тил илиминде колдонуу мүмкүнчүлүктөрү, андан алынган натыйжалардын тил илиминин өнүгүшүнө тийгизген таасирлери, тил илиминин тармактарына тиешелүүлүгүн негиздейт [Ибрагимов, 1964].

Кыргыз тилинин грамматикалык өзгөчөлүктөрүн математикалык жол менен текшерүүгө терең багыт алган окумуштуу зат атоочтун жөндөмө категориясынын статистикалык маалыматтары боюнча [Ибрагимов, 1966], кийин жалпы зат атооч сөздөрдүн колдонулуш жыштыгын аныктоого аракеттенди [Ибрагимов, 1969].

Сөздөрдүн морфологиялык өзгөчөлүктөрүн машиналык жол менен териштирүү тилдин жалпы грамматикалык табиятын аныктоого мүмкүн эместигин өзгөчө белгилөө менен автор өзүнүн кийинки эмгектеринде кыргыз

тилинин синтаксистик байланыш жасоо жолдорун [Ибрагимов, 1972], сөз түркүмдөрү менен сүйлөм мүчөлөрүнүн ортосундагы байланыш, катыш маселелерин комплекстүү кароо шартынын илимий жактан негиздүүлүгүн далилдеп, аларды ыктымалдуулук теориясынын параметри менен изилдөөдөн так жыйынтыктар алынарын айкындайт [Ибрагимов, 1975].

1975-жылы «Кыргыз тилиндеги синтаксистик катыштарына структура-ыктымалдык анализ жана алардын кепте функцияланышы» деген темада кандидаттык диссертациянын жакталышы ошол кезде кыргыз тили үчүн гана эмес, жалпы Совет тил илиминде колдонмо тил илиминин өнүгүшү жана калыптанышына зор салым болду [Ибрагимов, 1975].

Структура-ыктымалдык анализ функционалдык класс-терминине байланыштуу жүргүзүлүп, ал сөздөрдүн морфологиялык класстары жана алардын сүйлөмдө аткарган функциялары аркылуу аныкталат. Изилдөөдө модель, схема жана коэффициент сыяктуу математикалык формулалар кеңири пайдаланып, автор изилдөөдөн алынган натыйжалар катары төмөнкүлөрдү сунуштайт: «Теориялык планда тилди сыпаттоонун жаңы методун колдонуп, ал морфологиялык жана синтаксистик мүнөздөрдү чагылдырган формалдык класстарды негиздөөнүн жардамы аркылуу жүзөгө ашырды. Натыйжада кыргыз тилинин синтаксистик кубулуштарынын жаңы мүнөздөмөлөрү, айкындалды; статистикалык планда синтаксистик байланыштардын түрдүү типтерин айкындалып жана класстарга бөлүштүрүү боюнча колдонулган жаңы эсептөөлөр колдонулуп, ал 5 матрицанын пайда болушуна алып келди, айкаштык жана позициялык конфигурациялардын жыштыгын бөлүштүрүү бекемделди жана катыштык мүнөздөмөлөр такталды; практикалык планда алынган натыйжалар тилдик классификация жүргүзүүнүн теориясында жана тилди окутуунун методикасында колдонууга шарт түздү» [Ибрагимов, 1975, 18].

Кыргыз тил илиминде структуралык, колдонмо жана математикалык лингвистика адистиги боюнча алгачкы диссертациялык иш Р.Г.Пиотровскийдин келечектүү шакирттеринин бири Т.Садыков тарабынан жакталды [Садыков, 1982]. Ишке даярдык көрүүдө изилдөөчү алгач «Манас» эпосунун жыштык сөздүгүн түзүүгө негизделген ЭЭМде кыргыз тексттерин иштетүүнүн квантативдик-лингвистикалык структурасын изилдеп чыгып [Садыков, 1979], кийинки эмгегинде лингвистикалык автоматта окулган түрк сөз формаларын синтездөө иш-чараларын, жасалма интеллекттин лингвистикалык проблемаларын чечүүгө аракеттенгенин белгилесек болот [Садыков, 1980].

Кыргыз тилинин атоочтук сөз түркүмдөрүнүн регулярдуу формаларын жаратуучу (*порождающий*) механизмдер жана морфологияны моделдештирген системага арналган илимий изилдөө иши кийинки түрк морфемаларынын түпкүрлүк (*глубинная*) жана сырткы (*поверхностная*) структураларын аныктоого шарт түздү [Садыков, 1980; 1981]. Түрк сөз формаларын машиналык анализдөө жана синтездөө маселелери А.А.Пиотровская, М.А.Махмудов жана А.Бабаларовдор менен бирдикте

иштелип, Кишиневдогу жана Минсктеги илимий конференцияларда илимий чөйрөнүн талкуусуна коюлган [Садыков, 1978; 1982].

Т.Садыков түрк сөз формаларынын лингвистикалык автоматтын түзүүнү максат кылуу менен, аны чечүүдө төмөнкү маселелерди күн тартибине койду: морфемалардын мааниси жана формасы менен мүмкүн болгон шайкештиктерин эсептөө жолу, морфемалардын типтерин бөлүп көрсөтүү, баштапкы аффикстердин мүмкүн болгон формаларынын толук тизмесин аныктап, аларга карата оптималдуу методиканы тандап алуу, талдоого алынган морфеманын маанисин билдирүү үчүн сингармониялык каражаттарга эсеп жүргүзүп, контексттин типтерин айкындоо, белгиленген контекстте морфеманын реалдуу формасын аныкташ үчүн алардын моделдерин түзүп чыгуу, түзүлгөн моделдерди ЭЭМде кайрадан иштетип чыгуу (*реализация*) [Садыков, 1975,]. Түрк (*кыргыз*) атоочтук морфологиясын моделдештирүү жагдайында Т.Садыков тарабынан алынган негизги жыйынтыктар: тизмелик көз карашты пайдалануу менен кыргыз атоочторунун бардык формалары толугу менен сыпатталган; автоматтык грамматика жана аларды регулярдуу билдирүү формалары түзүлгөн; сөз формаларынын ичиндеги грамматикалык контексттердин мүмкүн болгон типтери бөлүнүп көрсөтүлгөн; морфема жана графемалардын айкалышында үндүүлөр гармониясынын 21 тип-механизмдери айкындалган; белгиленген графемалык-оператордук контекстте морфемалардын реалдуу формасынын өнүктүрмө модели түзүлгөн; Ес-ЭЭМде моделдердин тууралыгы жана экспланатордугу текшерүүдөн өтүп, ал кыргыз тилиндеги материалдарга башка түрк тилдеринин материалдарын салыштыруу жолу менен далилдеген. Бул текшерүү оң натыйжаларга жетишүүгө өбөлгө түзгөн [Садыков, 1975].

Түрк тилинде колдонмо лингвистика багыты боюнча изилдөөлөр өткөн кылымдын 90-жылдарында гана жандана баштады. 80-жылдардын аягында түрк тилинин материалдарында колдонмо тил илиминин башат маселелери чагылдырылды. Аталган изилдөөлөр Байрав Сухейла, Южел Тахсин, Кожаман Ахмет, Кемаль Офлазер, Мехмет Кара, Эфрасияп Гемалмаз, Тахир Тахироглу, Мелтем Турхан Йөндем, Ялчын Чеби тарабынан жүргүзүлдү.

Иште лексикография, анын тарыхы, кыргызча сөздүктөрдүн хронологиялык тизмеси, тарыхый сөздүктөр, Осмон түркчөсү сөздүктөрү тууралуу кеңири маалымат камтылды.

Кыргызча сөздүк түзүүгө Б.Ө.Орузбаева, В.В.Радлов, К.К.Юдахин, М.Н.Богданова, Х.Карасаев, Ю.Яншансин, Ж.Шүкүров, П.С.Панков, К.Тыныстанов, А.А.Алдашов, А.Жапанов, К.Карасаев, К.Сейдакматов, Г.Ж.Жамшитова, Б.А.Лунин, А.Үмөталиев, А.Стамбеков, Д.Чолпонкулов, С.Ж.Карабаева, А.А.Акматалиев жана анын жетекчилиги астында иштеген илимпоздордун тобу кыргыз лексикографиясына чоң салым кошту.

«КОМПЬЮТЕРДИК ЛЕКСИКОГРАФИЯНЫН НЕГИЗДЕРИ» деп аталган экинчи бапта сөздүк, лексикография, лексикографиянын тарыхы, тарыхый сөздүктөр, кыргызча сөздүктөр боюнча жазылган эмгектер, сөздүктөрдү классификациялоо, бөлүштүрүү маселеси, сөздүктөрдүн түрлөрү

жана өзгөчөлүктөрү, таанымал электрондук сөздүктөр жана котормо системалары, кыргызча электрондук сөздүктөрдүн үлгүлөрү, электрондук сөздүктөрдүн мүмкүнчүлүктөрү боюнча маалымат берилди.

Лексикография – бул сөздүк түзүүнүн теориясы жана практикасын изилдеген прикладдык лингвистиканын бир тармагы.

Компьютердик лексикографиянын негизги багыттары: 1. Компьютердик каражаттарды колдонуп, тексттерден автоматтык түрдө сөздүктөрдү (*жыштык, терминологиялык, конкорданс ж.б.*) түзүү; 2. NLP-системалары (*Natural Language Processing*) үчүн компьютердик сөздүктөрдү түзүүнүн теориялык жана практикалык аспектилери. 3. Салттык сөздүктөрдүн машиналык варианттагы сөздүк үлгүлөрүн түзүү.

Кыргыз тилинин лексикалык, семантикалык жана функционалдык маалымат базасын даярдоодон мурун бул системалардын изилдөөгө алынышы ишибизге бир топ салым кошот. Электрондук сөздүктүн маалымат базасы үчүн жалгыз гана электрондук сөздүктү изилдеп коюу жетишсиз. Анткени айрым электрондук сөздүктөрдөгү өзгөчөлүктөр башкасында болбой калган учурлар кездешет. Муну жоюунун бирден-бир жолу таанымал электрондук сөздүктөр Google котормочу (*Google translate*); котормочу Биң (*Bing translator*); Вавилон (*Babylon*); ABBYY тили (*ABBYY Lingvo*); добуштуу сөздүк (*Sesli Sözlük*); он-лайн чоң түркчө сөздүк (*Online Büyük Türkçe Sözlük*); кыргызча электрондук сөздүктөрдүн үлгүлөрү; инкорекс тилчи; Стардикт менен котормо системаларын жалпылай түрдө изилдеп, алардын ар биринин жетишкендиктерин таап, жаңы түзүлө турган электрондук сөздүккө ошол өзгөчөлүктөрдү киргизүү болуп эсептелет.

Кыргыз тилинин сөз байлыгынын лексикалык, семантикалык жана функционалдык маалымат базасынын көп багыттуу түзүлө турганын эске алсак, бул маалымат базасынын электрондук чөйрөнүн бардык тармактарында колдонула турган касиетке ээ болушу шарт. Бул маалымат базасы интернеттен да орун алат. Анткени, интернет чөйрөсүндөгү сөздүк бардык интернет колдонуучуларына бирдей мүмкүнчүлүк түзүп бергендей эле, кыргыз тилинин эл аралык аренадагы кадыр-баркын арттырат. Изилдөөбүздүн жыйынтыгында толук кандуу кыргызча он-лайн сөздүгү жок экени аныкталды. Мындай кемчиликти жоюу үчүн тилчилер менен программисттер бирге кыргыз тилинин он-лайн сөздүктөрүн түзүшү керек. Андай сөздүктү даярдоонун алдында интернет чөйрөсүндө башка тилдер үчүн иштеп жаткан он-лайн сөздүктөрдүн типтери жакшылап аныкталып, алардын артыкчылыктары менен кемчиликтерин билгенден кийин гана иш алып баруу кажет. Мына ушул себеп менен биз да өз кезегинде он-лайн сөздүктөрдүн ичинен экөөнө токтололу.

Он-лайн сөздүк түрк тилинин бардык тармактарынын сөз байлыгын камтып, аны интернет чөйрөсүндө орток маалымат базасында сунуштайт. Анын базасынан жазма тилдин сөз байлыгы менен катар, илим-билим, искусство жана спортко таандык терминдерди, жер-суу аттарын, энчилүү аттарды, диалекттердеги сөздөрдү, фразеологизмдерди табууга болот. Он-

лайндагы көлөмдүү түркчө сөздүк (*Online Büyük Türkçe Sözlük*) системасы аркылуу маанисин үйрөнүүнү каалаган сөздү тармагын, стилин белгилебей туруп эле алар тууралуу кеңири маалыматтарга ээ боло аласыз.

Жыйынтыктап айтканда, электрондук сөздүктөр изилдөөчүлөр, колдонуучулар үчүн абдан пайдалуу. Дүйнө жүзүндө техниканын абалын жана мындан кийинки келечегин ойлосо, болочок китеп, гезит-журнал, жазуу булактарынын баары интернет дүйнөсүнөн орун аларын байкайбыз.

Үчүнчү бап «**Лексикографиялык маалымат базасы**» деп аталып, анда маалымат базасы (*Database*), программалоо жана алгоритм, лексикографиялык маалымат базасынын компьютердик лингвистикадагы орду тууралуу сөз болуп, кыргыз тилинин сөз байлыгынын лексикалык, морфологиялык, семантикалык жана функционалдык маалымат базасын түзүүнүн жолдору көрсөтүлүп схемалар, формалар жана алгоритмдер түзүлдү.

Кыргыз тилинин сөз байлыгынын формалык, категориялык, семантикалык жана функционалдык маалымат базаларын түзүүчү программа даярдалды.

Маалымат базасы – бул маалыматтардын топтому. Анда маалыматтар түрдүү формаларга салынып сакталат. Баарыбызга белгилүү бул формага кылдаттык менен баа берип, так маалыматтарга ыкчам жетишип, өзгөртө алабыз.

Маалымат базасы аркылуу аткара турган иш-аракеттерибиз негизги беш бөлүктөн турат: 1. Маалымат кошуу; 2. Маалымат көрүү; 3. Маалымат табуу (*ылгоо*); 4. Маалымат өзгөртүү (*оңдоп-түзөө*); 5. Маалымат өчүрүү.

Маалымат базаларынын жогорку функцияларын арттырууда программалоо тилдеринин жана алгоритмдердин орду чоң. Программалоо тилдеринин бири менен алгоритм колдонбой түзүлгөн маалымат базасындагы маалыматтар чийки бойдон калат. Ушундан улам, маалымат базасын байытып, функциясын арттырып, функционалдык түргө алып келип, максаттарыбыз үчүн каалагандай колдоно турган программалоо тилдеринин жана алгоритмдердин болушу шарт.

Программалоо – бул компьютердин жабдыктарынын кандайча иштей турганын, буйруктарды, сөздөрдү, арифметикалык амалдарды жөнгө сала турган нерселердин жыйындысы. Ал «программалоо тили» деп аталган компьютердик тил аркылуу ишке ашат. Азыркы учурда андай программалоо тили катары *Java*, *C#* сыяктуу жогорку деңгээлдеги тилдер менен бирге, *C*, *assembly* жана кээде машина тили да колдонулат. Программалоонун функцияларына караганда, ал чыныгы жашоону чагылдырып турат. Кандайдыр бир программаны жазуу же маселени чечүү үчүн анын чыныгы жашоодо болуп жатканын элестетүү керек. Мында буйруктар менен коддор жөн гана каражат катары эсептелет. Чечим чыгарууда логикага ылайык кандайча жана кайсы методду колдонуп жасоо экенин билүү зарыл. Ал эми бул нерсе алгоритмдердин жардамы аркылуу жүзөгө ашат.

Алгоритм – кандайдыр бир маселени чечүү үчүн колдонулган метод. Бардык программалоо тилдеринин негизинде алгоритм жатат. Алгоритмсиз

чыныгы маанисинде программа түзүү мүмкүн эмес. Алгоритмдер үчүн негизги үч компонент колдонулат.

1. Коэффициенттер (Variable): Сөз=X; Мүчө=Y; Сөз + мүчө = натыйжа

Бул мисалдан байкалгандай, алгоритм түзүүдө коэффициенттерди аныктап алабыз. Маалымат базасы түзүлүп жатканда, ага кошулган ар бир сөз же мүчө коэффициенттердин бирин туюнтат.

2. Алгоритм. Маселени чечүүгө керектелген иреттүү, логикалык кадамдардын баарын алгоритм деп атайбыз. Кандайдыр бир алгоритм үчүн төмөнкүлөр сөзсүз аткарылат: • Ар бир кадам өтө так болуп, тобокелге салууга жол берилбейт; • Белгилүү санга жеткенден кийин, алгоритм жыйынтыкталат; • Алгоритмдерди салыштыруу үчүн бардык мүмкүнчүлүктөр алдын-ала эске алынат; • Алгоритмдердеги айрым кадамдар орун алмаштырылгыдай деңгээлде ийкем болот. Ошону менен бирге ордун которсо болбой турган алгоритмдерди да эстөө керек.

3. Диаграмма. Диаграмма – бул коэффициенттерди колдонуунун натыйжасында пайда болгон алгоритмдердин формалык жактан чагылдырылышы. Ал алгоритмдерди сызыктар, квадраттар, тик бурчтар ж.б. геометриялык фигуралардын негизинде көрсөтөт. Программалоодон мурда алгоритм түзүп жатканда катанын деңгээлин төмөнкү даражага түшүрүүгө жана аны жоюуга көмөктөшөт.

Программаны кандай программалаштырсак, үтүр-чекитине чейин биз айткандарды жасайт. Башкача айтканда, компьютерге эмне берсек, ошону алабыз. Эми бул мисалдагы программалоо логикасын сөздүк маалымат базасындагы маалыматтарды колдонуп, амалдарды аткаралы.

Кыргыз тилинин грамматикасындагы бардык формаларга алгоритмдер жана диаграммалар түзүлүшү талап кылынат. Кандайдыр бир тилдин алгоритмдери ошол тилдин табиятын мыкты билген кишилер тарабынан, б.а. тилчи-окумуштуулар тарабынан түзүлөт. Антпесе, туура эмес түзүлүп калган алгоритм бир категориядагы натыйжалардын бардыгынын туура эмес чыгышына себепкер болот. Мисалы, кандайдыр бир сөзгө кайсы учурларда алгач жак мүчө, кайсы учурларда чак мүчө жалганары ж.б. көптөгөн амалдар алгоритмдер менен диаграммаларды түзүп жатканда туура системалаштырууну талап кылат.

Сөздүк маалымат базасы жана алгоритми даярдалган кандайдыр бир тил бардык керектөөлөрдү канааттандырат. Андан кийинки иш-аракетте программалоону мыкты жүргүзүү гана калат. Ал эми программалоо алгоритмдердин коддоштурулушу дегенди түшүндүрөт. Програмалоо иш-аракетин жүргүзүү учурунда компьютер адистеринен кеңеш алып туруу зарыл. Бир киши тилди, тилдин эрежелерин жана программалоо тилин мыкты билсе, өз керегине жараша каалагандай программа түзө алат. Тилекке каршы, эки тармактын бирин гана билген кишиден андай натыйжаны күтүү кыйын. Ушундан улам, аталган эки тармак жакшы программа түзүү үчүн тыгыз байланышта болушу зарыл.

Компьютердик лингвистиканын эң негизги керектөөсү маалыматтык базаларга багытталат. Керектөөлөргө ылайык маалымат базалары түзүлбөй туруп, компьютердик лингвистика жана электрондук лексикография тууралуу сөз кылуу мүмкүн эмес. Ушундан улам, эң башкы максат маалымат базаларын түзүү экенин аңдайбыз.

Кыргыз лексикографиясынын маалымат базасын түзүүнүн негизги төрт багыты: **лексикалык** маалымат базасы; **морфологиялык** (*зат атооч, сын атооч, тактооч, ат атооч ж.б.*) маалымат базасы; **тармактык** жана **семантикалык** маалымат базасын түзүү; **функционалдык** маалымат базасы тууралуу кеңири сөз болуп, алдыбызга коюлган максатка жетүүгө аракет кылынды.

Диссертациялык ишибизде тил бирдиктеринин ар биринин өзгөчөлүктөрү өз-өзүнчө маалымат базаларында иштелгендей эле, бардыгын бир жерге топтоштурууга да мүмкүн болот. Анткени тилди жалпысынан бирдиктүү кароо үчүн бардык өзгөчөлүктөрдү бир программага камтуу компьютердик лингвистиканын ишин жеңилдетет. Эми бул маалымат базасы эмнелерди камтый турганы, кандай өзгөчөлүктөргө ээ болушу керектиги, кандайча даярдала турганы жана маалыматтарды киргизүү кайсы принциптердин негизинде ишке ашырыла турганы тууралуу сөз кылабыз.

Сөздөрдүн лексикалык касиеттери маалымат базасынан коддор аркылуу киргизилет. Бул абал кийин сөздөргө же маалымат базасына өзгөртүүлөрдү, түзөтүүлөрдү жүргүзүүгө жана маалыматтардын негизинде жаңы алгоритмдерди түзүүгө жагымдуу шарт түзүп берет.

Бир тилдин лексикографиялык маалымат базасы түзүлүп жатканда, эң маанилүү баскычтарынын бири сөз категорияларын аныктоо болуп саналат. Бул тармак лексикографиялык жактан да, компьютердик лингвистика жагынан да чоң мааниге ээ. Мисалы, кандайдыр бир иш-аракеттин алкагында маалымат базасындагы атооч сөздөргө автоматтык түрдө барыш жөндөмөсүнүн мүчөсүн жалгоо зарылчылыгы жаралса, система базадагы сөздөрдүн категорияларын изилдеп, алардын ичиндеги атооч сөз катары киргизилген сөздөрдүн баарына эрежелерге ылайык биз талап кылган мүчөнү кошуп чыгат да, натыйжаларды экранга сунуштайт. Сөздөрдүн үстүндө мындай иш-аракеттерди чексиз түрдө жүргүзүүгө болот. Андыктан сөздөрдү маалымат базасына киргизүүдөн мурун, өтө кылдаттык менен алардын категориялары аныкталат жана сөздөрдү кошуу терезеси ошого жараша даярдалат.

Маалымат базасына сөз кошуп жатканда, сөздөрдүн *зат атооч, сын атооч, тактооч, ат атооч, этиш ж.б.* сөз түркүмдүк касиеттерине маани берип тим болбостон, бир эле сөз түркүмүнө кирген конкреттүү сөздөрдүн деталдуу маалыматтарын берүү керек. Мисалы, сөздүк базага киргизиле турган сөз зат атооч болсо, анын жалпы же энчилүү ат (*адам, жаныбар, географиялык ж.б.*) экени белгиленет. Ушундан улам, сөз түркүмдөрүнүн баары боюнча аларга мүнөздүү категориялар сөз киргизүү формасына камтылат.

Бул илимий иште сөз түркүмдөрүнүн ар бири колго алынып, сөздүк базада алган орду, өзгөчөлүктөрү аныкталды жана ар биринин алгоритми өзүнчө схема түрүндө көрсөтүлдү.

Кыргыз тилинин сөз байлыгы боюнча маалымат базасын түзүүдөгү эн маанилүү тармактардын бири катары морфологиялык маалымат базасы сөз түркүмдөрүнүн негизинде толугу менен эске алынып, сөздүк статьялардын өзгөчөлүктөрү бааланып, анын негизинде ар бир сөз түркүмү боюнча схемалары чыгарылды. Пайда болгон схемалардын негизинде базага сөз кошуу үчүн сөз кошуу формалары да түзүлдү. Бул формалар аркылуу сөздөр базага кошулуп жатканда, кайсы сөз түркүмүнө таандык экени жана андагы өзгөчөлүктөр эске алынат. Кыргыз тилинин сөз байлыгы үчүн электрондук чөйрөдө түзүлө турган маалымат базасы компьютердик лингвистика менен электрондук лексикографиянын өнүгүшүнө зор салым кошот.

Бир тилдин лексикографиялык маалымат базасы түзүлүп, муну менен катар маалымат базасындагы сөздөрдүн өзгөчөлүктөрү толугу менен камтылат. Ал базада сөздөрдүн тармактары да белгиленет. Тармактык өзгөчөлүк сөздөрдү тизмектөө, семантизациялоо үчүн чоң мааниге ээ. Тармактык өзгөчөлүктөр сөздүн конкреттүү түрдө кайсы тармакка (*биология, физика, химия, география, астрономия, медицина, экономика, социология ж.б.*) таандык экенин көрсөтөт. Мында бир сөздүн бир эле учурда эки же андан ашык тармакта колдонулуп колдонулбагандыгына өзгөчө көңүл бурулуп, маалымат базасына ошол касиеттери боюнча киргизилет.

Сөздөрдүн семантизацияланышы үчүн сөз тармактарынын орду өзгөчө маанилүү. Ушундан улам, маалымат базасында сөздөрдүн маанилери же сөз колдонулган тармактар эске алынып семантизацияланат. Мисалы, «суу» сөзүн семантизациялап жатканда, жалпы маанисинен тышкары химия тармагына да киргизилет. Тармактык өзгөчөлүктөрү белгиленип түзүлгөн маалымат базасын кийин лингвистикалык жана электрондук лексикография жаатында колдонуу бир топ жеңилдиктерди түзөт. Анын негизинде мыкты деңгээлде түзүлгөн маалымат базасы аркылуу сөздөрдүн тизмесин түзүп, тармактар боюнча жаңы сөздүктөрдү даярдоо мүмкүнчүлүгүнө ээ боло алабыз.

Кандайдыр бир маалымат базасына кошуп жатканда, сөздөрдүн кайсы тармакка тиешелүү экени белгиленет: *автоматика; адабият; айыл чарбасы; астрономия; биология; география; грамматика; дин; зоология; искусство; кино; курулуш; лингвистика; математика; медицина; музыка; социология; спорт; тамак-аш; театр; телевидение; технология; туризм; укук; физика; философия; химия; экономика; жалпы ж.б.*

Жогорку тармактардан тышкары, керектөөлөргө жараша көптөгөн тармактарды кошууга болот. Ушул негизде даярдалган «сөздүк-тармак тандоо формасы» төмөнкүдөй болот.

Сөздөрдү маалымат базасына кошуп жатканда, алардын синоним, антоним жана жакын маанилери сөзсүз түрдө эске алынат. Мындай түрдөгү маалымат базасы аркылуу функционалдык иш-аракеттерди жүргүзүүгө болот. Маалымат базасы толугу менен даярдалып бүткөндөн кийин, «бардык

синоним сөздөрдү тизмекте», «бардык антоним сөздөрдү тизмекте» деп буйрук бере алабыз. Алынган натыйжаларды жөн гана керектөөлөрүбүз үчүн колдонуп тим болбостон, жаңы типтеги түрдүү сөздүктөрдү түзсө болот. Мындан тышкары, маалымат базасынан кандайдыр бир сөз издегенде, ал сөздүн маанилери менен катар, синоними, антоними бар сөздөрдү билүүгө жана аларды башка да иш-аракеттерге колдонууга шарт түзүлөт.

Маалымат базасына сөз киргизип жатканда, омоним сөздөрдү белгилөөнүн зарылчылыгы жок. Себеби сөздөр маалымат базасына кошулгандан кийин, омонимдерди программанын жардамы аркылуу автоматтык түрдө тизмектөөгө болот.

Маалымат базасына кошула турган сөздөрдүн тармактык түшүндүрмөсүн, маалымат базасына кошула турган сөздөрдүн тармактык түшүндүрмөсүн же башка тилдеги эквивалентин билүүнү кааласак, аларды маалымат базасына киргизебиз. Пайда болгон маалымат базасын тил үйрөтүүдө, автоматташтырылган котормо системаларында, он-лайн сөздүктөрдү түзүүдө ж.б. тармактарда эркин колдонуу үчүн сөздөрдүн түшүндүрмөлөрүн жана чет тилдеги маанилерин маалымат базасына киргизилишине же маалымат базасында иштей тургандай абалга алып келиниши зарыл.

Функционалдык маалымат базасын түзүү менен, сөздөрдү жалгыз гана бир максат эмес, көптөгөн кызматтар, тармактар үчүн да колдонууга жетишебиз. Азыр бул тармакта сөздөрдүн кайсы өзгөчөлүктөрүнүн орун алышына жана алардын кайсы тармактарда колдонулушу мүмкүн экенине токтололу.

Сөздөрдүн этимологиясы. Сөздөрдү базага кошуп жатып, сөздөрдүн форма жана маанилеринен тышкары сөздөрдүн төл сөз же өздөштүрүлгөн сөз экенин белгилейбиз. Бул лексикографиялык жана лингвистикалык иш-чаралар жалпыга пайда алып келет. Сөздөрдүн этимологиялары маалымат базаларына киргизилгенде, түрдүү функцияларда колдонууга өбөлгө түзүлөт. Алсак, кыргызчага арап тилинен кирген сын атоочтор же этиштер боюнча изилдөөнү каалаган тилчи-окумуштуу бир нече секунддун ичинде маалымат базасындагы сөздөрдүн жалпысы боюнча маалымат алууга жетишет. Же болбосо, кыргыз тилиндеги бардык орусча же арапча сөздөр тизмектелип, алардын статистикалары чыгарылып, жаңы сөздүктөрдү түзүү иштери жолго коюлат. Мындан тышкары, маалымат базасындагы кандайдыр бир сөздү изилдегибиз келсе, ал сөздүн түшүндүрмөсү, категориясы, тармагы ж.б. өзгөчөлүктөрү менен бирге анын кайсы тилден кыргызчага өздөштүрүлгөнү көрсөтүлөт. Ушуга окшош типтеги ондогон функция боюнча маалымат базасы даярдалышы ыктымал.

Сөздүн уңгусу. Маалымат базасына кошула турган сөздөр жөнөкөй, татаал жана туундуу сөздөрдөн турат. Булардын ичинен туунду сөздөрдүн уңгулары конкреттүү түрдө белгиленет. Натыйжада, бир уңгудан чыккан сөздөрдү тизмектөө, бир сөздүн кайсы мүчөлөрдү алганы, сөздөрдүн кандайча пайда болгону ж.б. көптөгөн маалыматтарды автоматтык түрдө бөлүп карай

алабыз. Мыкты сөздүк базасын түзүүнү кааласак, сөздөрдүн уңгулары да белгиленет. Ошондо гана маалымат базасынын функциясы артып, колдонуу чөйрөсү кеңеет.

Сөздүн уңгусун белгилегенден кийин, ал уңгунун формасы белгиленет. Маалымат базасына кошулган бир сөздүн уңгусу башка сөз түркүмүнө таандык болуп калышы мүмкүн. Мисалы, «бактылуу» деген сын атоочтун уңгусу «бакыт» деген зат атооч сөз. Мындай сөздөр маалымат базасына кошулуп жатканда, деталдуу түрдө иштелип чыкса, кийин уңгусу зат атооч болсо да, туундусу этиш сөздөрдү; уңгусу этиш болсо да, туундусу сын атоочко туура келген этиштерди ж.б. категориялардын үстүнөн чыпкалоо иштерин жүргүзсө болот. Уңгу сөздөрдүн мүчө алган формаларын, кайсы сөз түркүмүнөн оошконун, мүчөлөрдүн функцияларын, статистикаларын билип алабыз.

к, п тыбыштарынын өзгөрүүсү (аккомодация). Аягы *к, п* тыбыштары менен бүткөн сөздөргө үндүүдөн башталган мүчө жалганса, алар *г, б* тыбыштарына өтүп айтылышы боюнча жазылат: *көк + ыш = көгүш, бак + ар = багар, чап + уу = чабуу, китеп + ы = китеби, бак + ында = багында, топ + ым = тобум* ж.б. Тууранды сөздөрдө бул тыбыштар жумшарбайт: *такылда, чакылда, бапылда, тапылда* ж.б.

Сөздүк базасына сөздөрдү кошуп жатканда, аягы *к, п* тыбыштары менен бүткөн сөздөрдүн кайсыларында *к/г, б/п* өзгөрүүлөр болуп болбошу белгиленет. Анткени маалымат базасы тексттердин анализин жүргүзүп жатканда, сөздөргө айрым мүчөлөр жалганса, тыбыштык өзгөрүүлөрдү автоматтык түрдө жүргүзгөндөй болот. Мисалы, маалымат базасындагы бардык атооч сөздөргө таандык мүчө жалгаш керек деп алалы. Бул учурда компьютерде жасала турган иш-аракеттин формуласын төмөнкүдөй түзөбүз:

| Жактар | атооч | Таандык мүчөлөр |
|-----------|-------|-----------------|
| Менин | | -(Ы)м |
| Сенин | | -(Ы)ң |
| Сиздин | | -(Ы)ңЫз |
| Анын | | -Ы / -сЫ |
| Биздин | | -(Ы)ңЫз |
| Силердин | | -(Ы)ңАр |
| Сиздердин | | -(Ы)ңЫздАр |
| Алардын | | -Ы / -сЫ |

Алынган маалыматтардын негизинде компьютердик программа маалыматтык базадагы атоочторду иргеп, алардын аягына эрежелерге ылайык мүчөлөрдү жалгайт. Эгер биз маалымат базасына сөздөрдү кошуп жатканда, кайсы сөздөрдөн кийин *к/г, б/п* өзгөрүшү бар экенин көрсөтпөсөк, төмөнкү таблицадагыдай туура эмес натыйжаларды берип калат. Үлгү сөз – «китеп»

| Жактар | Туура эмес натыйжа (к/г, б/п базага киргизилбесе) | Туура натыйжа (к/г, б/п базага киргизилсе) |
|--------|--|---|
| Менин | китепим | китебим |

| | | |
|-----------|--------------|--------------|
| Сенин | китепиң | китебиң |
| Сиздин | китепиңиз | китебиңиз |
| Анын | китепи | китеби |
| Биздин | китепибиз | китебибиз |
| Силердин | китепиңер | китебиңер |
| Сиздердин | китепиңиздер | китебиңиздер |
| Алардын | китепи | китеби |

Башкача айтканда, компьютер *к* жана *п* тыбыштары менен аяктаган сөздөргө үндүүдөн башталган мүчө жалганганда, анализ жасабай туруп эле тиешелүү мүчөлөрдү сөздөргө жалгап, туура эмес натыйжаларды көрсөтөт.

Мындай көйгөйлөргө жол бербеш үчүн маалымат базасынын иш-аракеттерине өзгөчө көңүл бурабыз.

Элизия. Аягы үндүү менен бүткөн сөзгө үндүү менен башталган мүчө уланганда, уңгунун аягындагы же мүчөнүн башындагы үндүүлөрдүн бири сыйлыгышып түшүп калат: *алты+оо=алтоо, жети+өө=жетөө, башта + оо = баштоо, иште+өө=иштөө, же+ыш=жеш, ата+ым=атам* ж.б. [Акунова А.Р., 2001, 63-б]

Мындай өзгөрүүлөргө дуушар болгон сөздөрдү маалымат базасына кошуп жатканда, төмөнкүдөй форма колдонобуз.

Элизия

☐ бар
 ☒ жок

Редукция. Басым түшпөгөн муундагы үндүүлөрдүн басандап, угулуш сапатынын өзгөрүшү, түшүп калуу абалына чейин жетиши. Мындай кубулуш айрыкча кууш *[ы], [и], [у], [ү]* тыбыштарына тиешелүү. Алсак, *айыл+ым=айлым, кийим+ы=кийми, эрин+ы=эрди, моюн+у=мойну, бакыт+ы=бакты, орун+ы=орду, ууру+ла=уурда, жумуш+а=жумша, оюн+а=ойно* ж.б. [Акунова А. Р., 2001, 63-б]

Бул топко кирген сөздөрдүн ичинен сөздүк статьяга кирбей тургандар, б.а. сөз жасоочу мүчөлөр эмес, сөз өзгөртүүчү мүчөлөр жалганганда гана өзгөргөнү белгиленет. Анткени сөз жасоочу мүчөлөр аркылуу өзгөргөндөр ансыз деле сөздүк статья катары маалымат базасынан орун алат. Мисалы, *кыйын-кыйноо, жумуш-жумша, жыйын-жыйна* ж.б. сөздөрдөгү редукция кубулушун маалымат базасында белгилөөнүн кажети жок.

Сөз өзгөртүүчү мүчө алганда, редукция кубулушу жүргөн сөздөр маалымат базасында белгиленет. Мисалы, *айыл+ым=айлым, эрин+ы=эрди, моюн+у=мойну, бакыт+ы=бакты, орун+ы=орду* ж.б.

Муну эске албасак, аткарылган иш-аракеттерде көптөгөн каталар, кемчиликтер чыгышы мүмкүн. Ушундан улам, сөздөрдү базага кошуп жатканда редукциясы бар же жок экенин төмөнкүдөй форма аркылуу тактап алууга болот.

Редукция

☐ бар
 ☒ жок

Жогоруда ар бири боюнча изилдөө жүргүзүп, ал өзгөчөлүктөрдүн маалымат базасына киргизүү менен аякталат.

Корутунду бөлүмүндө, компьютердик лингвистика жана электрондук

лексикографиянын азыркы күндүн талабына толук жооп берип, керектөөлөрүбүздү канааттандыра турган деңгээлге көтөрүү үчүн төмөнкү иш-аракеттер максатка ылайык аткарылды:

- *Лингвистиканын, лексикографиянын кайсы тармагы боюнча изилдөө жүргүзсөк, ошол тармакка шайкеш келген тармактык керектөөлөрдү аныктап, маалымат базасына киргизүү;*

- *Катталган маалыматтарды иштетүүдө каалаган натыйжаларга жетишүү үчүн программалоо тилдеринин ичинен керектүүлөрүн колдонуп, тиешелүү жагдайларда программист адистерден кеңеш алып туруу;*

- *Маалыматтарды иликтөө жана издеген маалыматтарга жетишүү үчүн программалоо тилдеринен бизге керектүүсүн тандап алабыз. Анткени көптөгөн программалоо тилдеринин иштөө функциялары бири-бирин кайталагандыктан, алардын айрымдары гана биздин ишибизди жеңилдетет. Жетишилген натыйжалардын кайсы тармактарда колдонула тургандыгына токтолобуз. Интернет аркылуу иш алып бара турган система түзүп жаткан болсок, интернеттеги программалоо тилдеринин бирин (asp, aspNet, php, java, perl, cfm, xml ж.б.) тандап колдонобуз.*

- *Жүргүзүлө турган лингвистикалык изилдөөлөрдө материалдык муктаждыктар эмес, илимий талаптардын болушуна көңүл буруу. Себеби каржылык кызыкчылыктан улам жүргүзүлгөн иш-аракеттер шашылыш түрдө анчейин маани берилбей жасалгандыктан, системада кемчиликтерди пайда кылып, илимий мааниде алгылыктуу нерсеге жетишүүдөн бизди оолактатып салат.*

- *Тил тынымсыз өнүгүп өзгөрүп турган нерсе болгондуктан, түзүлө турган системаларда буга ылайык убакыт менен өзгөртүлүп тура турган мүмкүнчүлүктөрдү кароо;*

- *Балким, булардын ичинен эң маанилүүсү – университеттерде, илимий изилдөө мекемелеринде ж.б. тармактарда жекече изилдөө топтору түзүлүп, алардын ортосунда маалымат алмашуу жана кыска убакыт ичинде маалыматтарды табуу мүмкүнчүлүктөрүн түзүп берүү;*

Изилдөөнүн негизинде ээ болгон маалыматтардын көмөгү менен программисттер жана тилчи-окумуштуулар түзө турган программалар төмөнкүдөй алгылыктуу жыйынтыктарга алып келет:

- *Кыргыз тили колдонулуп жаткан электрондук чөйрөнү көзөмөлдөп, андагы тилдин тууралыгы текшерилет жана программаны колдонуу учурунда айрым эскертүүлөр аркылуу колдонуучулардын кыргызчаны туура колдонуусун камсыз кылат.*

- *Өркүндөтүлгөн сөз маалымат базасын түзгөндөн кийин, аны колдонууга сунуштоо менен, ар түрдүү программалардын даярдалышына шарт түзүп, тилге байланышы бар кайсы гана тармакта болбосун алардын пайдаланылышын ишке ашырат.*

- *Кыргыз тилинин жана программалоо тилдеринин логикасы колдонулуп даярдала турган түрдүү программалар аркылуу тил үйрөтүү жана үйрөнүү процессин алда канча ыкчам абалга алып келет.*

- Маалымат базасындагы сөздөрдү колдонуп, тексттин анализин жүргүзүп, тармактарын белгилеп, сөз жыштыктарын аныктап, жетишилген маалыматтар аркылуу текстти баалоого шарт түзөт.

- Туура алгоритмдерди түзүү менен, компьютерлердин үлгү сүйлөмдөрдү түзүшүн, туура эмес сүйлөмдөрдү, сөздөрдү жана мүчөлөрдү таап чыгып, оңдойт.

- Автоматизациялык системалардын үн буйруктарды кабыл алуудагы кемчиликтерин жоё турган маалымат базаларын өркүндөтөт.

- Диаграммалары, алгоритмдери даярдалган кандайдыр бир тил формуласы кадимки эле математикалык маселе чыгаруунун так өзүн чагылдырып турат. Формуланы билсеңиз, аны колдонуп, математикалык маселени чече аласыз. Тилдин эрежелери үчүн алгоритминиз даяр болсо, бардык электрондук чөйрөлөрдө өз каалаган натыйжаңызды аласыз.

- Иште сунушталган сөз маалымат базасы, алгоритми жана диаграммалар коштолуп даярдалган программалар электрондук чөйрөдө, негизинен, үч чоң функцияны аткарат: анализ жүргүзүү, оңдоо-түзөө жана жаратуу.

Анализ жүргүзүү. Даярдала турган программалар компьютердеги же интернеттеги, б.а. электрондук чөйрөдөгү кандай гана жазуу болбосун алардагы кыргыз тилинин туура колдонулуп жатканын анализдеп, статистика жүргүзүп, бизди керектүү маалыматтар менен камсыздап, өзүнө ишенген учурларда бизден сурабай туруп эле өзү автоматтык түрдө текстке оңдоп-түздөө киргизет.

Оңдоо-түздөө. Жогоруда айрым үлгүлөрүн берген анализдин натыйжасында пайда болгон ката жана кемчиликтерди программа автоматтык түрдө өзү оңдойт. Ал эми оңдоого күчү жетпей калган учурларда программа колдонуучусуна кабар берет.

Жаратуу. Кыргыз тилинин бүтүндөй структуралык, семантикалык, синтаксистик жана грамматикалык эрежелерин программалоонун натыйжасында пайда болгон система кийинчерээк көптөгөн тармактарда жаңы нерселерди пайда кылуучу программа катары колдонулат. Кыскасы, компьютердик программалоо тилдеринин кайсы гана түрүн колдонсо да, кыргыз тилинин эрежелери алгоритмдер жана функциялар аркылуу компьютерлерге үйрөтүлүп, компьютердик базаларга орнотулат. Кыргыз тилинин эрежелерин жана анын сөз байлыгын өздөштүргөн система кыргызча сүйлөмдөрдү куруп, белгилүү эрежелердин негизинде жазма жана оозеки түрдө суралган суроолорго анализ жүргүзүп жооп берет.

Лингвистика, лексикография жана компьютер тармактарын бириктирүү үчүн алардын ар бирин мыкты билүү керектиги турган иш. Азыркы учурда программалоо тилдери менен лингвистика жаатында жогорку билимге каныккан адис-изилдөөчүлөрдүн саны чектелүү болгондуктан, бул нерсе көйгөй жаратууда.

Маселени чечип, компьютердик лингвистиканы мындан ары карай өркүндөтүү үчүн эки тармакта иш алып баргандар өз ара кызматташууга даяр болуулары кажет, же бир тармакта эмгектенгендер экинчи тармакты да өз

кесибиндей мыкты өздөштүрүүгө милдеттүү. Ошондо гана болуп көрбөгөндөй натыйжаларга жетишилип, анын негизинде сансыз программаларды түзүп, тажрыйбалар артып, компьютердик лингвистика тармагында койгон максаттарга жетишүүгө болот. Андан кийин робот технологиялары, автоматизациялык системалар, синхрондук котормо, текст анализи ж.б. тармактар компьютердик лингвистика жараткан бай тажрыйбадан пайдалана баштайт.

Ал үчүн университеттердин лингвистикага байланыштуу бөлүмдөрүндө программалоонун негиздери, алгоритмдер жана диаграммалар тууралуу сабактар берилип, азыркы күндүн талаптарына ылайык технологияларды кандайча колдонуу керектиги үйрөтүлүүгө тийиш.

Азыркы кылымды технологиянын кылымы деп эсептесек, анда бул чындыкты далилдеп, тилибизди технологияга ылайыктап, аны эффективдүү колдонууга аракет кылуу – замандын талабы.

Диссертациянын негизги мазмуну төмөнкү макалаларда чагылдырылды:

1. «Кыргыз лексикографиясынын өнүгүш этаптары», Наука и новые технологии ISSN 1026-9045 No 3, Бишкек 2010, 271-275- беттер.

2. «Сөздүктөрдү басмага даярдоо принциптери жана классификациялоо маселелери», Известия ВУЗов, ISSN 1026-9045 № 4, Бишкек 2010, 270-276-бб..

3. «Бүгүнкү күндөгү лексикографиянын өнүгүшү жана электрондук сөздүктөр», Наука и новые технологии, ISSN 1026-9045 № 1, Бишкек 2011, 280-284-бб..

4. «Dilbilim, Sözlükbilim ve Bilgisayar» Dil ve Edebiyat Araştırmaları Dergisi The Journal of Turkish Language and Literature Studies, Year 02, issue 03, Summe 2011, ISSN 1308 - 5069, İstanbul, 41-53- бб..

5. «Компьютерге кыргыз тилин үйрөтүү жана кыргыз тилинин сөз байлыгы боюнча маалымат базасын түзүү принциптери», НАУКА И НОВЫЕ ТЕХНОЛОГИИ, ISSN 1026-9045 No 4, Бишкек 2011, 278-284- бб..

6. «Popüler Çeviri Sistemleri, Özellikleri ve Eksiklikleri», Социальные и гуманитарные науки, ISSN 1026-9045 No 1-2, Бишкек 2011, 174-182- бб..

7. «Компьютердик лексикография», Социальные и гуманитарные науки, ISSN 1026-9045 No 3-4, Бишкек 2011, 180-185- бб..

8. «Кыргыз тилинин сөз байлыгынын категориялык (сөз түркүмдүк) маалымат базасын түзүү», К.Карасаев атындагы Бишкек гуманитардык университетинин «Кыргыз тили жана адабияты» журналы, Бишкек 2012.

Керим Сарыгүлдүн «**Кыргыз тилинин лексикографиялык маалымат базасын түзүү маселеси**» деген темада 10.02.01 – кыргыз тили адистиги боюнча филология илимдеринин кандидаты окумуштуулук даражасын изденип алуу үчүн жазылган диссертациясынын

РЕЗЮМЕСИ

Багыттоочу сөздөр: кыргыз тили, маалымат базасы, лексикография, алгоритм, диаграмма, коэффициент, программа, компьютер, интернет

Изилдөө объектиси: кыргыз тилинин лексикографиясы

Изилдөөнүн максаты: кыргыз тилинин лексикографиялык маалымат базасын түзүү.

Иште колдонулган методдор: Изилдөөдө математикалык методдор, салыштырма усулу, сыпаттама, сурамжылоо, байкоо жүргүзүү, системалаштыруу, моделдештирүү сыяктуу методдор колдонулду.

Алынган жыйынтыктар жана анын жаңылыгы: Кандайдыр бир тилдин керектөөлөрүнө жана технологиялык мүмкүнчүлүктөрүнө жооп берүүсү, түрдүү тармактарда натыйжалуу колдонулушу тилчилер менен колдонуучулардын көз карашынан алганда, аткарыла турган иш-аракеттердин үч баскычтан турары анык болот:

1. Өркүндөтүлгөн сөздүк маалымат базасын түзүү;
2. Тилдик эрежелердин негизинде алгоритмдерди жана диаграммаларды түзүү;
3. Программаларды түзүү.

Сөздүк маалымат базасы түзүлүп жатканда, сөздөрдүн кайсы өзгөчөлүктөрү эске алынары, алардын маалымат базасында кандайча сакталары, тилдик алгоритмдер менен диаграммалардын лингвистикалык мааниси эске алынуу менен, үлгүлөр камтылып, бул тармакта иш алып бара турган изилдөөчүлөргө багыттар көрсөтүлдү.

Колдонуу чөйрөсү: Изилдөөнүн натыйжалары, тыянактары, диссертациянын материалдары лексикология, лексикография, семантика, стилистика ж.б. дисциплиналары боюнча лекция өтүүдө, окуу китептери менен окуу куралдарын жазууда, сөздүктөрдү түзүүдө, котормодо баалуу материал болуп кызмат кыла алат.

Бул илимий эмгекти изилдөөчүлөр жогорку окуу жайларында аталган темада атайын тандоо курстарын окутууда колдонсо болот.

К. Сарыгүл

РЕЗЮМЕ

к диссертационному исследованию Керим Сарыгүл на тему: «Проблемы создания лексикографической базы кыргызского языка» на соискание ученой степени кандидата филологических наук по специальности 10.02.01 – кыргызский язык.

Ключевые слова: кыргызский язык, база данных, лексикография, алгоритм, диаграмма, коэффициент, программа, компьютер, интернет

Объект исследования: лексикография кыргызского языка

Цель исследования: Целью диссертационной работы является создание базы данных кыргызской лексикографии

Методы исследования: В работе использованы математический метод, сравнительный (контрастивный) методы, описательный, опрос, наблюдение, систематизирование, моделирование

Полученные результаты исследования и их научная новизна: Научная новизна заключается в том, что подобное исследование проводится впервые.

Обеспечение потребностей и технологических возможностей того или иного языка, функциональность в разных областях, по мнению языковедов, состоит из трех ступеней:

1. Создание развитой базы данных;
2. На основе языковых правил создание алгоритмов и диаграмм;
3. Создание программ.

Были указаны пути исследования ученым создающих лексикографических баз данных; были выявлены особенности слов при введении в базу данных, и их особенность в сохранении в базе данных; лингвистическое значение языковых алгоритмов и диаграмм; были приведены примеры.

Область применения: Материалы и результаты работы могут быть применены в области лексикологии, семантики и стилистики языка; в научных исследованиях; в подготовке научно-методической литературы; в проведении лекций в высших учебных заведениях; в составлении словарей и т.д.



RESUME

Kerim Sarygul

Topic: « **The problems of creating the lexicographic base of the Kyrgyz language** »

The thesis is submitted to confer the scholarly degree of Candidate of Philological Sciences

Specialty 10.02.01 - kyrgyz language

Key words: *Kyrgyz language, database, lexicography, algorithm, diagram, coefficient, program, computer, internet*

Object of research: lexicography of the Kyrgyz language

Goal of dissertation: The aim of the thesis is to creation of a database of Kyrgyz lexicography

Methods of research: mathematical method, comparative (contrastive) methods, descriptive, survey, observation, systematization, modeling

Results of the research and scientific novelty: The scientific novelty lies in the fact that such a study is conducted for the first time.

Ensuring the needs and technological capabilities of a language, functionality in different areas, according to linguists, consists of three stages:

1. Creation of a developed database;
2. Based on language rules, the creation of algorithms and diagrams;
3. Creating programs.

The ways of research to scientists creating lexicographic databases were indicated; The peculiarities of words were introduced when entering the database, and their peculiarity in saving in the database; linguistic significance of language algorithms and diagrams; examples were given.

Range of application: Materials and results can be used in research, preparation of scientific and methodical literature, lectures at universities, compiling dictionaries, etc. in the field of linguistics, lexicology, semantics and stylistics.



