

**БИШКЕКСКИЙ ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ**  
**ИМЕНИ К. КАРАСАЕВА**  
**КЫРГЫЗСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**  
**ИМЕНИ И. АРАБАЕВА**  
**ИНСТИТУТ ЯЗЫКА И ЛИТЕРАТУРЫ**  
**ИМЕНИ Ч. АЙТМАТОВА ПРИ НАН КР**

Д 10.17.548 Диссертационный совет

На правах рукописи  
УДК: 811.512.154(575.2)(04)

**САРЫГУЛ КЕРИМ**

**ВОПРОСЫ СОЗДАНИЯ ЛЕКСИКОГРАФИЧЕСКОЙ БАЗЫ**  
**КЫРГЫЗСКОГО ЯЗЫКА**

по специальности 10.02.01 – кыргызский язык

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата филологических наук

**Бишкек – 2018**

Работа выполнена в институте языка и литературы им.Ч.Айтматова при НАН КР.

**Научный руководитель:**

доктор филологических наук, профессор <b>Токоев Таалайбек Токтосунович</b>
---

**Официальные оппоненты:** доктор филологических наук, профессор  
**Садыков Талполот Садыкович**

кандидат филологических наук, доцент  
**Абдыракова Гульмира Шарапидинова**

**Ведущая организация:** Иссык-Кульский Государственный  
Университет им. К. Тыныстанова, кафедра  
кыргызского языкознания.  
Адрес: 722200, г. Каракол, ул. Ж.  
Абдырахманова, 103.

Защита состоится «8» июня 2018 года в 10:00 на заседании специализированного диссертационного совета Д.10.17.548 по защите диссертаций на соискание ученой степени доктора (кандидата) филологических наук при бишкекском гуманитарном университете им.К.Карасаева, кыргызском государственном университете им.И.Арабаева, Институте языка и литературы им.Ч.Айтматова при НАН КР по адресу: 720044, г. Бишкек, пр. Мира 27.

Автореферат разослан «7» мая 2018 г.

Ученый секретарь  
Диссертационного совета, к.ф.н., доцент

Каратаева С.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования** обусловлена тем, что такие направления языкознания как общее языкознание, лексикография, фонология и т.д. нуждаются в достижении кратчайшие сроки желаемых результатов в сфере электронной технологии.

Сегодня в мире создана тысяча компьютерных программ. В будущем это явление будет развиваться в ускоренном темпе. Тем не менее, очень мало из существующих программ можно будет использовать в языкознании. В последние годы участилась практика внедрения в электронную систему словарей и текстов.

Несмотря на развитие этого направления, в электронную систему в неизменном виде внедряются сырые тексты и словари.

Это явление препятствует развитию лингвистики, лексикографии, текстологии, фонологии, семантики и т.д.

Электронный вариант информации экономит для нас время во время поиска информации.

Электронизация информации огромное значение имеют во всех направлениях языкознания.

Электронные данные должны быть полными, а также легкими для использования. Внесение в базу данных нижеперечисленных деталей во время создания электронного словаря значительно облегчат работу при анализе, редакции, создании текста: *является какой частью речи (существительное, прилагательное, наречие, глагол и т.д.), не является ли слово исключением, корень слова, является ли данное слово заимствованным словом, список антонимов и синонимов и т.д.* Эти слова и их определения находятся в одной таблице, а также каждый признак должен разместиться в отдельной ячейке. Поскольку компьютеры могут анализировать только в строгом соответствии с инструкциями.

С точки зрения лингвистики и лексикографии, в разработке широкоформатных баз данных кыргызского языка, очень важен соответствие с современными технологиями его диаграмм и алгоритмов.

**Связь темы диссертации с крупными научными программами, основными научно-исследовательскими работами, проводимыми научными учреждениями.** Диссертация выполнена в рамках научно-исследовательских работ, государственных программ, проектов, а также является индивидуальной научной работой.

**Цель и задачи исследования.** Целью исследования является разработка новых баз данных лексикографии кыргызского языка.

**Для достижения цели диссертационной работы были поставлены следующие задачи:**

1. Разработать диаграмм и алгоритмов широкоформатных баз данных кыргызского языка, соответствующих современным технологиям.

2. Определить степень важности диаграмм и алгоритмов для языкознания (особенно для компьютерной лингвистики), лексикографии (особенно для электронной лингвистики).

3. Дать инструкцию использования данных с помощью языков программирования.

4. Показать возможность использования полученных результатов не только в языкознании, но и в других научных направлениях.

**Научная новизна диссертации заключается в том,** что подобное исследование проводится впервые.

По мнению лингвистов, обеспечение потребностей лексикографии того или иного языка с использованием технологических возможностей состоит из трех ступеней:

1. Создание развитую базу данных;
2. Создание алгоритмов и диаграмм на основе языковых правил;
3. Создание программ.

Были указаны пути исследования учеными создающих лексикографических баз данных; были выявлены особенности слов при введении в базу данных; указаны лингвистические значения языковых алгоритмов и диаграмм; закреплены примерами.

**Теоретическая и практическая значимость работы.** Основные положения и выводы диссертационного исследования могут быть использованы в широких научных кругах. На кыргызском языке впервые написана подобная диссертационная работа. Все эти признаки являются научной новизной. Новую программу можно использовать во всех направлениях языка. Программа имеет основных трех функций (*анализ, редактирование, доработка*): лексикография / электронная лексикография; лингвистика / компьютерная лингвистика; обучение к языкам; анализ текста; спряжение/предложение/составление текста; перевод; автоматические системы; робототехнологии; интернет и т.д.

**Основные положения и результаты исследования, выносимые на защиту.** Для достижения колоссальных результатов программисты и лингвисты должны работать сообща. В результате подобных совместных работ программа:

1. Берет под свой контроль электронную систему кыргызского языка; проверяет правильность языков электронной системы; обеспечивает правильное использование кыргызского языка;

2. Создание много программ с помощью базы данных продвинутого словаря. Обеспечение использование этих данных не только в языкознании, но и в других направлениях науки связанные с языкознанием.

3. Использование логики языка программирования и ускорение процесса обучения и изучения кыргызского языка.

4. Анализировать текст с использованием слов баз данных. Определяется направление текста, уточняется частотность слов, текст оценивается с использованием этих информации.

5. Обеспечивается составление предложений компьютером на основе правильно работающих алгоритмов.

6. Расширить возможность функций автоматизированных систем при принятии голосовых приказов.

Вышеперечисленные функции являются основными. Кроме этого можно создать сотни похожих программ, а также развивать уже существующих программ. Учитывая схожесть структур всех тюркских языков, изменив только некоторых особенностей программ данного языка, можно использовать и для других родственных языков.

**Личный вклад соискателя** состоит в непосредственном проведении в проведении всех исследований, апробации результатов.

Соискателем создана база данных лексикографии кыргызского языка.

Определены отличия между простыми и развитыми базами данных. Исследована поле их использования.

Созданы базы данных и алгоритмы с целью широкого использования кыргызского языка в эпохе технологии, даны инструкции ученым по использованию их.

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

**Апробации результатов диссертации.** Результаты диссертации опубликованы в 8-и статьях.

**Структура и объем работы диссертации.** Диссертационная работа состоит из 251 страниц основного текста, 14 схем, 10 форм, 47 рисунков и 1 программы.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** диссертации содержатся следующие ключевые разделы: актуальность темы диссертации, цель, задачи, научная новизна, теоретическая и практическая значимость, основные положения, результаты исследования, использованные методы, апробация, структура и объем диссертации.

**Первая глава** диссертации называется: **«СТАНОВЛЕНИЕ ПРИКЛАДНОЙ ЛИНГВИСТИКИ И ЕЕ РАЗВИТИЕ»**. Эта глава состоит из пяти частей. В ней дается общая информация о прикладной лингвистике, нынешнем состоянии, развитии прикладной лингвистики в тюркологии, обзор научных литератур прикладной лингвистики тюркологии, кыргызского языкознания и турецкого языкознания, их история и основные особенности лексикографии.

Прикладная лингвистика отличается своим широким и всеобъемлющим процессом формирования и развития. Основная часть его развития и формирования приходится к XX-XXI вв. Следует отметить, по традиции его основу составляют письменность (*графика*), методика преподавания родного, иностранного языков и лексикографические данные языка. Если обратим внимание на развитие языка, то можно заметить, что на более поздних стадиях развития переводоведения научная литература по дешифровке, орфографии, транслитерации, терминологии были направлены на решение практических вопросов языка и рассмотрены в тесной связи с прикладной лингвистикой.

Бодуэн Дэ Куртенэ одним из первых поставил вопрос о точных методах в лингвистике, предлагал выделять слова на основе строгих процедур. Бодуэн Дэ Куртенэ широко используя числа, математические задачи в языкознании, утверждал, что языкознание должно быть намного ближе к точным наукам [Бодуэн де Куртенэ, 1963].

В настоящее время все шире развивается использование научных и технических терминов в разных направлениях. Как известно, но этот вопрос рассматривается за пределами лексикографии. В настоящее время в обществе очень популярно направление с широким использованием компьютеров автоматизированной лексикографии.

Прикладная лингвистика — сложное направление в языкознании тесно связанная с философией, психологией, физиологией, математикой, физикой, логикой, информатикой. Данное направление языкознания учёными называется по-разному: «Компьютерная лингвистика», «инженерная лингвистика», «автоматическая лингвистика», «счетная лингвистика», «квантативная лингвистика», «статистическая лингвистика».

Первые научные труды о прикладной лингвистике были написаны ещё в начале XIX века. Несмотря на это, только во второй половине XX века была сформирована его научно-теоретическая часть, которая позволила ей стать перспективной и самостоятельной наукой.

Сегодня исследовательское пространство этого языкознания очень

расширился. В начале 1957 года написан первый научный труд, который назывался «статистика языка». Основоположниками дескриптивного и статистического языкознания были американский лингвист Г.Глисон, русский языковед Р.Г.Пиотровский [Глисон,1959; Пиотровский, 1959].

60-х годов прошлого века прикладная лингвистика в тюркологии шагнула далеко вперед. Ленинградской школой лингвистики Советского Союза во главе Р.Г.Пиотровским, П.М.Алексеевым по направлению «статистика языка» были проведены значимые исследования. С момента создания «словаря языка Абая» в 1968-м году статистическая лингвистика Казакстана была подвержена более интенсивному развитию.

Прикладная лингвистика в Казакстане связана с именами К.Б.Бектаева, Д.А.Байтанаевой, А.К.Жубанова, а Узбекистана С.А.Ризаева, в Татарстане К.Р.Галиуллина, А.А.Тухватуллиной, Л.С.Шарипниязовой, Р.Г.Бухарева, Б.С.Сафиуллиной, Д.Ш.Сулейманова, в Азербайджане К.В.Велиевой, Д.Жабарлы, М.Махмудова, А.Фатуллаева и т.д.

Направление прикладной лингвистики кыргызского языка связано с открытием лаборатории экспериментальной фонетики при институте языка и литературы Академии наук КР. Этим институтом были проведены экспериментальные исследования звуковым системам кыргызского языка. Результаты вышесказанного исследования отражены в трудах Т.К.Ахматова, Ж.К.Сыдыкова, А.О.Орусбаева, Т.Чымбаева, Т.Садыкова.

Подготовка молодых ученых Кыргызстана в аспирантурах Москвы и Ленинграда в конце 70-х годов обеспечила в будущем компетентными кадрами. В трудах профессора С.Ибрагимова впервые отражены решения языковых вопросов математическим методом и закрепление их с теорией вероятности.

Одним из первых С. Ибрагимов предложил использовать математических методов в кыргызской лингвистике, а также считать влияние полученных результатов как развитие языкознания и причастность к языковым направлениям [Ибрагимов, 1964].

С.Ибрагимов хотел проверить особенностей грамматики кыргызского языка математическим способом. Статистическими данными категорий падежа существительного, позже можно заметить, как автор старался уточнить частотность всех существительных слов [Ибрагимов, 1969].

С.Ибрагимов утверждает невозможность определения характера грамматики языка с перечислением автоматическим путем морфологических особенностей слов. В следующих работах автор пишет о создании синтаксических связей кыргызского языка между частями речи и членов предложения. Он доказал, что рассмотрение взаимоотношений комплексно и исследование по теории вероятности дают точных результатов [Ибрагимов, 1975].

Защита кандидатской диссертации на тему «Структурно-вероятностный анализ синтаксиса кыргызского языка и их функции в речи» в 1975 году внес значительный вклад не только на развитие прикладной лингвистики кыргызского языка, но и общей Советской лингвистике

[Ибрагимов, 1975].

Структурно-вероятностный анализ проводится в связи с функциональной класс-терминами. Он определяется морфологическими классами слов и их функций в предложениях. В этом исследовании широко использованы такие математические формулы как модель, схема и коэффициент. Автор в качестве результатов исследования предложил следующее: «Были использованы новые методы в теоретическом описании языка. Это было реализовано с помощью обоснования формальных классов отражающих морфологических и синтаксических аспектов. В первую очередь, выявлены новые факты синтаксических явлений кыргызского языка; определены различные типы синтаксических связей в синтаксисе; использованы новые расчеты классификации. Это привело к появлению пяти матриц. Закреплены распределения соотношений частотности и позиционных конфигураций; обеспечены в теоретическом классификации языка и методики преподавания языков использование результатов полученных в практическом плане [Ибрагимов, 1975, 18].

Первая диссертация по направлению структурной, прикладной и математической лингвистики кыргызского языкознания была защищена учеником Р.Г. Пиотровского Т.Садыковым [Садыков, 1982]. В рамках диссертации он исследовал квантативно-лингвистическую систему текстов кыргызского языка, создавая частотный словарь эпоса «Манас» на компьютере [Садыков, 1979]. В следующих работах исследовал синтез турецких словоформ прочитанных лингвистическим автоматом, старался решить проблем искусственного интеллекта [Садыков, 1980; 1980].

Исследования порождающих механизмов и систем моделирования частей речи кыргызского языка дали информации о глубинных и поверхностных структурах турецкого языка [Садыков, 1980; 1981]. Вопросы автоматических анализов и синтезов словоформ турецкого языка разработаны совместно с А.А.Пиотровским, М.А.Махмудовым и А.Бабаларовым. Это широко обсуждалось на научных конференциях в Кишиневе и Минске [Садыков, 1978; 1982].

Т.Садыков с целью создания лингвистических автоматизаций словоформ турецкого языка выдвинул следующие вопросы: определение значений морфем с его формами показывая классификаций типов морфем; составление полных списков всевозможных форм начальных аффиксов; выбор оптимальной методики исследования; для определения значений морфем учесть особенность сингармонических средств; определение типов контекста, составить модель для реальных форм морфем в контексте; реализация составленных моделей на компьютере [Садыков, 1975, 3-4]. Проблема моделирования морфологии кыргызского языка при помощи электронной машины нашла отражение в монографии Т.Садыкова «Проблемы моделирования тюркской морфологии» (1987). Основными результатами исследования можно отметить следующее: с использованием списков описаны все формы частей речи; созданы автоматическая грамматика и их регулярная форма; классифицированы всевозможные типы



словоформ грамматических контекстов; определены 21 типов-механизмов гармонии гласных морфем и графем; составлены развивающие модели реальных форм морфем в графемо-операторском контексте; проверены правильность и экспланаторность моделей ЕС-ЭЭМ; это было доказано сопоставительным путем материалов кыргызского языка с другими турецкими языками; эта проверка способствовала достижению положительных результатов [Садыков, 1975, 15].

Можно отметить, что научные исследования прикладной лингвистики турецкого языка начались в 90-х годах прошлого века. Не трудно заметить, что первые проблемы прикладной лингвистики отражены в материалах турецкого языка в конце 80-х годов прошлого века.

Эти исследования связаны с именем Байрав Сухейла, Южел Тахсин, Кожаман Ахмет, Кемаль Офлазер, Мехмет Кара, Эфрасияп Гемалмаз, Тахир Тахироглу, Мелтем Турхан Йондем, Ялчын Челеби.

В диссертации широко рассмотрены лексикография, его история, хронологический список кыргызских словарей, исторические словари, словари Османско-турецкого.

Огромный вклад на создание кыргызских словарей внесли Б.О.Орузбаева, В.В.Радлов, К.К.Юдахин, М.Н.Богданова, Х.Карасаев, Ю.Яншансин, Ж.Шукуров, П.С.Панков, К.Тыныстанов, А.А.Алдашов, А.Жапанов, К.Карасаев, К.Сейдакматов, Г.Ж.Жамшитова, Б.А.Лунин, А.Умоталиев, А.Стамбеков, Д.Чолпонкулов, С.Ж.Карабаева, А.А.Акматалиев и т.д.

**Вторая глава** диссертации называется **«ОСНОВЫ КОМПЬЮТЕРНОЙ ЛЕКСИКОГРАФИИ»**. Она представляет собой обзор научных трудов о словарях, лексикографии, истории лексикографии, словарей кыргызского языка, вопросах классификации слов, разновидностях и особенностях словарей, популярности электронных словарей и систем перевода, примерах электронных кыргызских словарей, возможностях электронных словарей.

**Лексикография** — раздел языкознания, занимающийся вопросами составления словарей и их изучения; наука, изучающая семантическую структуру слова, особенности слов, их толкование. Лексикография – одно из направлений языкознания изучающая теорию и практику прикладной лингвистики.

Основные направления компьютерной лексикографии:

1. При помощи компьютерных средств создание автоматических словарей из текстов (частотные, теоретические, конкордокс и т.д.).
2. Теоретические и практические аспекты создания компьютерных словарей для NLP-систем (Natural language processing).
3. Создание электронных вариантов традиционных словарей.

Прежде создания формальных, семантических и функциональных баз данных кыргызского языка, исследование этих систем значительно облегчит нашу работу. Для создания баз данных электронных словарей недостаточно исследование только электронных словарей. Потому что, некоторые функции

электронных словарей не встречаются в других. Единственным способом решения этих проблем является одновременное исследование популярных электронных словарей и систем перевода, определение преимуществ каждого из них и внесение этих функций к новым электронным словарям.

Дана подробная информация и проведены исследования.

В настоящее время в мире из самых популярных систем перевода наиболее часто используемыми являются Google переводчик; Переводчик Bing; Вавилон; АBBYY; Большой голосовой словарь турецкого языка; Инкорекс; Стардикт.

Если учесть многосторонность создания формальных, семантических, и функциональных баз данных словарный запас кыргызского языка, то можно сказать что эта база данных может быть использована во всех аспектах электронного пространства. Она будет иметь своё достойное место и в интернете. Для решения этих проблем лингвисты и программисты должны работать совместно. При создании таких словарей нужно хорошо определить тип онлайн словарей других языков, рассмотреть их достоинства и недостатки.

По этой причине мы решили сконцентрироваться на двух онлайн-словарях.

Онлайн словарь турецкого языка охватывает все направления лексики. Его предлагает в общую базу данных в интернет пространстве. В его базе можно найти словарный запас письменного языка, научных терминов, терминов искусства, терминов спорта, топонимов, наименований вод, собственных имён, диалектов, фразеологизмов. Можно узнать значение слов Большого турецкого словаря, даже не зная направление и стиль слова.

В заключении можно сказать, современные технологии уже давно стали неотъемлемой составляющей нашей жизни. Словарь и лексикография будут существовать вместе с человеком и языком. Поэтому использование продвинутой технологии является потребностью наших дней.

**Третья глава** называется «**Лексикографическая база данных**». В этой главе дается информация о базе данных (*Database*), программировании и алгоритмах. Также значение базы данных лексики в компьютерной лингвистике, указаны пути, схемы, формы и алгоритмы формальных, категоричных, семантических и функциональных баз данных.

Подготовлены программы создания формальных, категорических, семантических и функциональных баз данных.

Информационная база представляет собой набор данных. В ней информации хранятся в различных формах. Как известно, мы можем оценить, получить быструю и точную информацию, также мы можем изменить её.

Действия базы данных состоят из пяти ступеней: 1. Добавление информации; 2. Обзор информации; 3. Поиск информации (сортировка); 4. Изменение информации (*редактирование*); 5. Уничтожение информации.

При совершенствовании функций баз данных имеют большое значение языки и алгоритмы программ. Информация без алгоритмов,

созданная с помощью одного из языков программирования в базе данных остаются незавершенными. Существование часто используемых языков программирования и алгоритмов являются обязательными для обогащения баз данных, а также улучшения функций, создания функциональных форм.

Программирование – это комплекс отладка и тестирование программы (*испытания программ*), анализ и постановка задачи, написание текстов программ, регулирование арифметических задач. Это реализуется на «языке программирования». В настоящее время, самыми известными языками программирования являются **Java, C# и C, assembly** иногда используется язык машины. По сравнению функциями программирования, она отражает реальную жизнь. Для создания каких-либо программ или для решений вопросов, нужно представить себе, будто все это происходит в реальной жизни. В этом случае приказы и коды рассматриваются только как средства. Самое главное, нужно знать использование какого метода является логичным для решения задач. Все это реализуется при помощи алгоритмов.

**Алгоритм** – это точно определенная инструкция, которую последовательно применяя к исходным данным, можно получить решение задачи. На основе всех языков программирования лежит алгоритм. Без алгоритма невозможно создать программу. Три основных компонентов используются для алгоритма.

**1. Коэффициенты (*Variable*):** Слово=X; Окончание=Y; Слово + окончание = результат

Как заметили в этом примере, при создании алгоритмов мы должны обязательно определить коэффициенты. Действительно, при создании базы данных, каждое слово или его окончание обозначают коэффициенты.

**2. Алгоритм.** Нужные порядковые и логические ступени при решении задач называются алгоритмами. Алгоритм должен обладать следующими свойствами:

- Каждый шаг должен осуществляться очень точно, без риска;
- После достижения определенной стадии, алгоритм завершается;
- Для сравнения алгоритмов всех возможностей должны быть рассмотрены заранее;
- Некоторые ступени алгоритмов должны быть гибкими, чтобы их можно было заменить. Также нужно учесть незаменимых алгоритмов.

**3. Диаграмма.** Диаграмма – это формальное отражение алгоритмов появившихся с использованием коэффициентов. Она показывает алгоритмов в виде линий, квадратов и других геометрических фигур. Помогает устранить ошибок во время создания алгоритмов в программировании.

Программа работает в зависимости от нашего программирования, т.е. компьютер возвращает то, что мы ему дали. Теперь мы постараемся решить задачу с использованием логики программирования информации словарной базы данных.

Требуется составление алгоритмов и диаграмм всем формам грамматики кыргызского языка. Алгоритмы какого-либо языка составляется учёными-языковедами знающими этот язык хорошо. В противном случае,

неправильно составленные алгоритмы будут причиной неверных решений при составлении алгоритмов и диаграмм. Систематизация задач должна быть правильной, например, какой слове первым добавляется окончания лица, а в каких окончания указывающие на время и т.д.

База данных и алгоритм словаря должны удовлетворять все потребности. После этого остается стадия создания хорошего программирования. Программирование означает кодирование алгоритмов. Во время программирования необходимо принимать советы от специалистов. Человек, хорошо знающий язык, правил языка и язык программирования, может составить необходимую программу. К сожалению, ожидать результатов от человека, который знает только один из них тяжело. Поэтому, для создания программы эти два направления должны разрабатываться одновременно.

Компьютерная лингвистика ориентирована на потребности базы данных. Невозможно говорить о компьютерной лингвистике и электронной лексикографии не составив необходимых баз данных. Учитывая этот момент, мы понимаем, что основной целью является создание баз данных.

Основные четыре направления базы данных кыргызской лексикографии: формальная база данных; категориальная база данных (*существительное, прилагательное, наречие, местоимение и т.д.*); создание базы данных по направлениям и семантических баз данных; функциональная база данных.

Как и каждая особенность размещены отдельно в базе данных, также будет возможно их сгруппирование. Потому что, рассмотрение языка как одно целое охватывая в одной программе намного облегчает работу компьютерной лингвистики. Теперь будем говорить об особенностях базы данных, методах создания и принципах реализации.

Формальные особенности слов ведутся в базу данных с помощью кодов. Такое положение может дать возможность создания новых алгоритмов на основе изменений базы данных, а также внести поправки и т.д.

Во время создания базы данных лексики того или иного языка, одним из важных действий является определение категорий слов. Это направление с точки зрения лексикографии и компьютерной лингвистики имеет огромное значение. Например, при необходимости к существительным в базе данных автоматически добавляется окончание дательного падежа. Система анализирует категорий слов в базе, добавляет необходимых окончаний по правилам к словам введенных в качестве существительных и результаты будут показаны на экране. Можно совершать таких действий безограничений. Поэтому прежде ввода слов в базу данных, нужно тщательно определить их категорию и окошка ввода слова должен создаваться в соответствии с этим.

При добавлении слов в базу данных нельзя ограничиться с информацией этой или той части речи: *существительное, прилагательное, наречие, глагол и т.д.* Нужно дать детализированную информацию о словах

входящих в одну часть речи. Например, если слово, введенное в базу данных, является существительным, то нужно обязательно отметить является ли оно нарицательным или именами собственными (*человек, животное и т.д.*). В итоге все качества частей речи будут включены в форму введения слов.

В этой диссертационной работе исследованы все части речи, их место в базе данных, определены особенности и указаны отдельно алгоритмы каждого из них.

Во время создания базы данных лексики кыргызского языка одним из самых важных направлений является категориальная база данных, основанная на частях речи. Также должны оцениваться словарные статьи, на основе этого отражены схемы частей речи. Специально созданы формы добавления слов по схемам. Во время добавлений слов в эту форму, указаны важность их особенностей и принадлежности к частям речи. Мы глубоко верим в то, что база данных лексики кыргызского языка созданная в электронном пространстве внесет огромный вклад на развитие компьютерной лингвистики и электронной лексикографии.

Если нужно создать базу данных лексики того или иного языка, вместе с этим отметить всех особенностей слов находящихся в базе данных, то должны указать направление слов в базе данных. В семантизации огромное значение имеет выявление направлений слов.

Точно указывается конкретный отрасль (*биология, химия, физика, география, астрономия, медицина, экономика, социология и т.д.*) того или иного слова. В этом случае нужно обратить внимание на то, относится ли слово только одному направлению или нескольким. В базу данных ведется и такая информация.

Для семантизации слов очень важно его направление. Поэтому слова семантизируются с учетом их значений в базе данных и направлений. Например, при семантизации слова «вода» можно указать кроме общего значения отношение этого слова и к химии.

Использование базы данных с указанием на направление в лингвистической и электронной лексикографии значительно облегчит нам задачу. На основе этого с помощью качественных баз данных можно составить список слов, а также дать возможность добавить новых слов по направлениям.

При добавлении информации в ту или иную базу данных, обязательно нужно указать направление слова: *автоматика; литература; сельское хозяйство; астрономия; биология; география; грамматика; религия; зоология; искусство; кино; строительство; лингвистика; математика; медицина; музыка; социология; спорт; пищевая инженерия; театр; телевидение; технология; туризм; право; физика; философия; химия; экономика и т.д.*

Помимо вышеуказанных направлений, по необходимости можно добавить многочисленных направлений. «Словарь – форма с выбором направлений» созданная с этим принципом выглядит следующим образом:

При добавлении слов в базу данных обязательно должны учитываться

их синонимы, антонимы и слова близкие по значению. При помощи такой базы данных можно реализовать функциональных работ. Когда база данных полностью подготовлена, можно дать следующие приказы: «создать список синонимов», «создать список всех антонимов». Можно не ограничиваться только с удовлетворением своих потребностей, но и создать словарей нового типа. Кроме этого, при поиске слова в базе данных вместе со значениями слов, дается список синонимов, антонимов. А также можно использовать для других действий.

При вводе слов в базу данных, нет необходимости определять омонимов. Когда слова будут добавлены в базу данных при помощи программы можно автоматически определить омонимов.

Если хотим узнать направление слов добавляемых в базу данных или эквивалент в других языках, то добавляем их в базу данных. Важно обратить внимание на определения слов, внесение значений в иностранном языке в базу данных для свободного использования баз данных в обучении к языкам, автоматизированных системах перевода, создание онлайн словарей и т.д.

На примере следующей формы можно дать эквиваленты на турецком и русском языках наряду с определением слова и в кыргызском языке.

Функциональную базу данных слов можно использовать не только с одной целью, а ещё во многих направлениях для разных действий. Теперь постараемся дать разъяснение в этом направлении какие особенности должны расположиться и возможность использования их.

Во время добавления слов в базу данных кроме форм и значений слов нужно обязательно отметить является ли данное слово заимствованным или исконным словом языка. Это принесёт пользу и в лексикографию, а также в лингвистику. Этимологию слов внесенных в баз данных можно использовать для разных функций. Например, ученый-языковед исследующий заимствованных прилагательных и глаголов с арабского на кыргызский может за считанные секунды найти нужную информацию в базе данных или получить сразу полный список всех русских и арабских слов, подсчитать их статистику, создать новых видов словарей. Кроме этого, если захотим исследовать то или иное слово вместе с определением, категорией, направлением и т.д. других особенностей, также будет указан с какого языка заимствована на кыргызский язык. Возможна подготовка базы данных с десятками функций подобного типа.

Слова внесенные в базу данных бывают *простыми*, *сложными* и *дериватами*. Среди них корень *дериватов* должен конкретно указан. В результате можно автоматически четко видеть *дериватов* из одного корня, какие окончания были добавлены той или иной слове, происхождение слов и т.д. Для достижений этих целей указание корней слов являются обязательными. Только в этом случае функция базы данных увеличивается, расширяется поле использования.

После фиксации корня слова, нужно определить форму корня. В базе данных корень слова может фиксироваться как другой часть речи. Например, корень прилагательного «счастливый» существительное «счастье». Если эти

особенности будут указаны при введении слов в базу данных, то потом легко можно фильтровать слов. Например, если корень слова является существительным, а дериват глаголом, или корень глагол, а дериват прилагательное и т.д. Можно подсчитать статистику форм исконных кыргызских слов с окончаниями, отношение к частям речи, функций окончаний и т.д.

Форма корня слова должна быть такой: изменение звуков *к, п* (аккомодация) – если словом заканчивающихся со звуком *к,п*, они меняются на звуки *г, б*: *көк + ыш = көгүш*, *бак + ар = багар*, *чап + уу = чабуу*, *китеп + ы = китеби*, *бак + ында = багында*, *топ + ым = тобум* и т.д.

Звуки в подражательных словах не меняются: *такылда, чакылда, бапылда, тапылда* и т.д.

Во время добавлений слов в базу данных, слова заканчивающиеся на буквы *к, п*, обязательно отмечаются в каких словах буквы меняются *к* на *г*, *б* на *п*, а в каких не меняются. Потому что при анализе текстов база данных должна сама автоматически менять окончания слов. Например, нужно добавить притяжательные окончания всем существительным.

В таком случае, давайте представим формул действий разработанных на компьютере.

Лицо	существительное	Притяжательные окончания
Менин (мой, моя, моё)		-(Ы)м
Сенин (твой, твоя, твоё)		-(Ы)ң
Сиздин (ваш, ваша, ваше)		-(Ы)ңЫз
Анын (его, её)		-Ы / -сЫ
Биздин (наш, наша, наше)		-(Ы)ңЫз
Силердин (Ваш, Ваша, Ваше)		-(Ы)ңАр
Сиздердин (Ваш, Ваша, Ваше)		-(Ы)ңЫздАр
Алардын (их)		-Ы / -сЫ

Программа базы данных фильтрует существительных, по правилам добавляет окончаний в конце слова. Если мы не укажем в каких ситуациях в конце слова *к* на *г*, *б* на *п* меняются, то мы можем получить неправильных результатов, как указано в нижеуказанной таблице.

Например: слово – «книга»

Лицо	Правильный результат ( <i>к/г, б/п</i> если не введен в базу)	Неправильный результат ( <i>к/г, б/п</i> если введен в базу)
Менин (мой, моя, моё)	китепим	китебим

Сенин (твой, твоя, твоё)	китепиң	китебиң
Сиздин (ваш, ваша, ваше)	китепиңиз	китебиңиз
Анын (его, её)	китепи	китеби
Биздин (наш, наша, наше)	китепибиз	китебибиз
Силердин (Ваш, Ваша, Ваше)	китепиңер	китебиңер
Сиздердин (Ваш, Ваша, Ваше)	китепиңиздер	китебиңиздер
Алардын (их)	китепи	китеби

Другими словами, при добавлении окончаний начинающихся на гласных к словам заканчивающихся на *к* и *п*, программа дала неправильных результатов. Чтобы предотвратить таких ошибок, нужно обратить особое внимание на действия базы данных.

Элизия – это отпадение звука в конце корня или в начале окончания: *алты+оо=алтоо*, *жети+өө=жетөө*, *башта + оо = баштоо*, *иште+өө=иштөө*, *же+ыш=жеиш*, *ата+ым=атам* и т.д. [Акунова А.Р., 2001, 63-с.].

В процессе добавления подвергающихся к таким изменениям слов в базу данных, нам следует использовать следующую форму:

Редукция – это оглушение или изменение гласных в безударных слогах. Такое явление особенно свойственны звукам [ы], [и], [у], [ү]. Например, *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* и т.д. [Акунова А. Р., 2001, 63-с.].

В словарную статью не включаются словообразующие окончания, а только изменяемые окончания слов. Поскольку изменяемые окончания слов с помощью словообразующих окончаний размещаются в базе данных как словарные статьи. Например, явление редукции в словах *кыйын-кыйноо*, *жумуш-жумша*, *жыйын-жыйна* и т.д. не обязательно отметить в базе данных.

При добавлении изменяемых окончаний слов в базе данных отмечаются слова подвергнутые к явлениям редукции. Например: *айыл+ым=айлым*, *эрин+ы=эрди*, *моюн+у=мойну*, *бакыт+ы=бакты*, *орун+ы=орду* и т.д.

Если не обратим на это внимание, то можем допустить много ошибок при выполнении действий. Поэтому при добавлении слов в базу данных, нужно уточнить о свойственности редукции.

В разделе **выводы** даны результаты исследования: проведены работы с целью поднятия уровня и обеспечения удовлетворений потребностей компьютерной лингвистики и электронной лексикографии. Указаны направления использования программ созданных программистами и



учёными-языковедами о функциях (*анализирование, редактирование и создание*) электронной системы с алгоритмами и диаграммами базы данных.

Изобретаемые программы должны анализировать правильность использования кыргызского языка в электронном пространстве, подсчитать статистику, обеспечить нас нужными информациями и при необходимости редактировать тексты автоматически.

Достаточно внести в программу функций программ анализирующий текст. В процессе дистанционного обучения к языкам, а также анализ текста компьютером самостоятельно, уведомление студента результатами в электронном пространстве. Если потребуется, чтобы студент написал сочинение, то после выполнения задачи студент отправит его нам через компьютер. В этом случае, компьютер может самостоятельно проанализировать текст, результат может отправить и студенту и преподавателю одновременно.

Программа не ограничивается только анализом текста. Кроме этого, он записывает и распознает голос. Программами обеспечиваются интеграция голоса к аппарату. Соответственно, оценив голосовой текст по правилам кыргызского языка, аппарат выявляет недостатки и ошибки.

Программа исправляет автоматически ошибки и недостатки, выявленных в результате анализа. А в случаях, когда программа не в силах исправить, то предупреждает пользователя об этом.

Мы можем с уверенностью сказать, что на первых порах эффективность таких программ бывает низко, но с течением времени, когда начнет широко использоваться, то недостатки исчерпаются и обеспечиваются точной информацией.

Для закрепления этого мнения можем привести пример стадий развития программ перевода компании Google. Если в первых порах переводы были очень смешными, то со временем значительно развивался и в настоящий момент достиг уровня, который может удовлетворить пользователей.

Созданная система в результате программирования структурных, семантических, синтаксических и грамматических правил кыргызского языка позже будет использована в качестве программы создающей новых вещей во многих направлениях. В независимости использования любым языком компьютерного программирования, правила кыргызского языка с помощью алгоритмов и функций вносятся к компьютерным базам. Система освоившая лексику и правила кыргызского языка может составить предложения, ответить устным и письменным вопросом по правилам. Это можно закрепить следующим примером: если приказать программе с алгоритмами правил кыргызского языка проспрягать в настоящем времени слово, то программа автоматически нам даст результаты по прежней логике. Для получения правильных результатов нужно обратить внимание на правильное введение данного глагола в настоящем времени заранее. Потому что для получения правильных результатов мы должны внести в компьютер заранее правил языка, закон сингармонизма, формы изменения глагола, явления элизии и

эпентези и т.д.

Только в этом случае система может показать нам желаемых результатов. Система, охватывающая все особенности языка, всегда будет показывать правильные ответы на основе формул и алгоритмов. В действительности, человек, изучающий иностранный язык, тоже проходит через этот путь. Если посмотреть с этой стороны, то существует сходство между человеком и компьютером. Но человек, изучающий иностранный язык для запоминания информации выполняет упражнения и старается запомнить их. Но компьютерам не свойственно забыть. Каждая операция осуществляется по этому принципу. Во время программирования системы или обучения человека к иностранным языкам допущение ошибок в грамматике приведут к плачевным результатам.

Для объединения лингвистики, лексикографии и компьютера нужно хорошо знать каждого из них. В настоящее время ограниченность подобных специалистов-исследователей порождает ряд проблем.

Чтобы решить эту проблему. Для развития компьютерной лингвистики исследователи ведущие работу в двух направлениях должны сотрудничать друг с другом. Учёные, которые специализировались только в одном направлении, должны освоить второе направление также хорошо, как свою профессию. Только в этом случае можно добиться беспрецедентных результатов, на основе этого создать безчисленных программ, добиться целей поставленных по направлению компьютерной лингвистики. После этого такие направления как робототехнологии, автоматизированные системы, синхронный перевод, анализ текста и т.д. могут воспользоваться богатым опытом компьютерной лингвистики.

Для этого в отделениях лингвистики при университетах должны вести уроки по «основе программирования, алгоритмов и диаграмм». В соответствии с требованиями сегодняшнего дня должны проводить лекции об использовании технологий.

Если мы считаем этот век эпохой технологии, то все мы должны доказать эту реальность, и попытаться эффективного применения технологий во всех направлениях лексикографии.

Мы считаем, что эта диссертационная работа достигла всех целей, которые были поставлены перед нами.

#### **Список опубликованных работ по теме диссертации.**

1. «Кыргыз лексикографиясынын өнүгүш этаптары», Наука и новые технологии ISSN 1026-9045 No 3, Бишкек 2010, 271-275- сс..
2. «Сөздүктөрдү басмага даярдоо принциптери жана классификациялоо маселелери», Известия ВУЗов, ISSN 1026-9045 № 4, Бишкек 2010, 270-276- сс..
3. «Бүгүнкү күндөгү лексикографиянын өнүгүшү жана электрондук сөздүктөр», Наука и новые технологии, ISSN 1026-9045 № 1, Бишкек 2011, 280-284-сс..
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- сс..

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- сс..

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

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

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 и т.д. 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.

## РЕЗЮМЕ

к диссертационному исследованию Керим Сарыгул на тему: **«Вопросы 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.