

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК КЫРГЫЗСКОЙ РЕСПУБЛИКИ
Институт физико-технических проблем и материаловедения им. Ж.Жээнбаева

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ КЫРГЫЗСКОЙ РЕСПУБЛИКИ
Кыргызский Государственный Технический Университет им. И. Раззакова
Кыргызско-Российский Славянский Университет им. Б. Ельцина

Диссертационный совет Д.05.11.034

На правах рукописи
УДК 681.5

Исраилова Нелла Амантаевна

**РАЗРАБОТКА АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ
ЛИНГВИСТИЧЕСКОГО ПЕРЕВОДА**

Специальность: 05.13.06-"Автоматизация и управление технологическими процессами и производствами"

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени
кандидата технических наук

Бишкек -2013

Работа выполнена в Кыргызском Государственном Техническом Университете им. И. Раззакова

Научный руководитель: к.т.н., доцент Шабданов М.А.

Официальные оппоненты: доктор технических наук, профессор Аккозиев И.А.

кандидат технических наук, с.н.с.
Исмаилов Ж.А.

Ведущая организация: Институт новых информационных технологий при КГУСТА им. Н.Исанова

Защита состоится «29» ноября 2013 года в 16.00 часов на заседании Межведомственного диссертационного совета Д.05.11.034 при Институте физико-технических проблем и материаловедения им. Ж. Жеенбаева НАН КР, КГТУ им. И. Раззакова, КРСУ им. Б. Ельцина по адресу: 720071, г.Бишкек, пр. Чуй, 265-а центральный корпус НАН КР.

С диссертацией можно ознакомиться в Центральной научной библиотеке НАН КР по адресу: 720071, г.Бишкек, пр. Чуй, 265-а.

Автореферат разослан «25» октября 2013 года.

Ученый секретарь Межведомственного диссертационного совета Д.05.11.034 кандидат физико-математических наук



В.В.Алиферов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Цель данной разработки продиктована отсутствием качественных программных продуктов, выполняющих лингвистический перевод в русско-кыргызском направлении. Бесспорную актуальность работы можно подчеркнуть Постановлением Правительства Кыргызской Республики от 5 марта 2013 года № 114 «О мерах по обучению государственных и муниципальных служащих государственному языку и переводу делопроизводства на государственный язык» в целях исполнения Закона Кыргызской Республики «О государственном языке Кыргызской Республики», постановления Жогорку Кенеша Кыргызской Республики «О комплексных мерах по реализации норм Закона Кыргызской Республики «О государственном языке Кыргызской Республики» от 22 апреля 2011 года № 370-V.

В данной диссертационной работе разработана автоматизированная система перевода текста официально-делового стиля с русского языка на кыргызский, реализованная с элементами экспертной системы. Такой путь представляется более эффективным при автоматизации технологического процесса лингвистического перевода. Так как все необходимые правила и данные, необходимые для осуществления перевода представляются в виде базы знаний, формируемых и корректируемых непосредственно лингвистом – экспертом на формализованном языке с помощью удобного и простого интерфейса системы. Это позволяет эксперту самому корректировать работу системы путем оценки качества полученного системой перевода и выполнения соответствующей её настройки.

Иначе говоря, разработанная автоматизированная система лингвистического перевода снабжена механизмами обучения и накопления знаний, что позволяет ей совершенствоваться с каждым сеансом перевода.

Связь темы диссертации с научными программами и научно-исследовательскими работами.

Работа была выполнена в Кыргызском Государственном техническом университете им. И. Раззакова в рамках НИР и госбюджетных проектов университета.

Цель и задачи исследования.

Целью исследования является разработка автоматизированной системы лингвистического перевода текста с русского языка на кыргызский.

Поставленная цель достигается решением следующих основных задач:

- разработка концептуальной модели системы;
- разработка базы знаний, построенной на формализованном языке в виде правил, схем и моделей, наиболее полно отражающих конструкции языка;
- разработка морфологического анализатора;
- разработка синтаксического анализатора;
- разработка алгоритмов обучения и оценки перевода;
- разработка алгоритма отладки процесса перевода.

Объектом исследования является технология процесса лингвистического перевода текста с русского языка на кыргызский.

Научная новизна диссертационной работы заключается в следующем:

- впервые разработана автоматизированная система перевода текста с русского языка на кыргызский, реализованная с элементами экспертной системы;
- разработана принципиально новая эффективная структура транслятора с оптимальным распределением функций между человеком-редактором и компьютером с целью повышения качества перевода и уменьшения затрат времени на перевод за счет обеспечения двух режимов работы: режима приобретения знаний и режима их применения в трансляции;

- разработана эффективная структура базы знаний системы, формируемая лингвистом в процессе перевода;
- определена оптимальность применения комбинированных алгоритмов перевода и их реализация.

Практическую значимость работы представляют:

- разработанный пакет прикладных программ для автоматизированного перевода текста в русско-кыргызском направлении;
- реализованные в работе комбинированные лингвистические алгоритмы;
- разработанные алгоритмы обучения системы и механизмы накопления знаний;

Результаты работы внедрены в работу отдела регламента и делопроизводства Аппарата Правительства КР и получены положительные отзывы от экспертов. Разработанная автоматизированная система защищена авторскими свидетельствами Кыргызпатента.

На защиту выносятся следующие основные положения:

1. Концептуальная схема функционирования автоматизированной системы лингвистического перевода;
2. Алгоритм функционирования морфологического анализатора;
3. Алгоритм функционирования синтаксического анализатора;
4. Алгоритм функционирования модуля отладки перевода;
5. Алгоритм обучения и приобретения знаний системы.

Личный вклад соискателя.

Все основные результаты, изложенные в диссертационной работе, получены лично соискателем.

Апробация работы.

Основные положения и результаты исследований по теме диссертации докладывались и обсуждались на следующих научных конференциях:

- Международной конференции "Телекоммуникационные и информационные технологии. Состояние и проблемы развития", Бишкек, 2001 г.
- Международной конференции, посвященной 20-летию кафедры ЭВМ КГТУ-ФПИ, Бишкек, 2002г.
- НТК молодых ученых "Инновации - вектор для молодежи", Бишкек, 2011

Публикации.

Основное содержание диссертации опубликовано в 13 работах. По материалам выполненных исследований получено 2 авторских свидетельства КР.

Структура и объем работы.

Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы и приложения. Работа содержит 136 страниц, в том числе 43 иллюстраций, 2 таблицы, а также 2 приложения. Список использованной литературы включает 99 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе проведен анализ состояния проблем автоматизации лингвистического перевода, приведены основные современные методы решения и алгоритмы, а также обозначена реальная актуальность разработки системы автоматизированного перевода текста в русско-кыргызском направлении.

В настоящее время разработано большое количество систем автоматического языкового перевода на различные пары или даже множества языков (англо-русские, русско-английские, русско-немецкие, русско-французские, и т.д.). Но, наряду с этим, можно отметить отсутствие такого разнообразия в русско-кыргызском или кыргызско-русском направлении. Более того, существующие программные продукты автоматизированного перевода на этой паре языков не имеют должного

уровня качества. Однако темпы компьютеризации общества набирают обороты и в нашей республике, появляется реальная потребность в системах автоматизированного перевода, именно, в русско-кыргызском направлении. В частности, необходимость перевода официальной документации с русского языка на кыргызский требует появления на рынке подобных программных продуктов с более качественной и гибкой системой функционирования.

Это позволило бы, может и не полного решения всех возложенных задач, но, по крайней мере, облегчило бы работу тех ведомств, которые занимаются непосредственным переводом официальной документации с русского языка на кыргызский.

Если учесть, что максимальная производительность труда переводчика составляет 4-5 авторских листов в месяц, то такая, например, система машинного перевода, как SYSTRAN переводит до 1000000 словоупотреблений в час, а это около 120 авторских листов. Из чего следует, что производительность систем машинного перевода превышает производительность человека-переводчика, порядка, в 20 раз. Эти количественные показатели свидетельствуют о преимуществе компьютерных систем. Однако качество такого перевода значительно уступает переводу, выполненному человеком. Тем не менее, этот факт должен только стимулировать процесс развития в сфере машинного перевода с переходом на более высокий качественный уровень.

Во второй главе выполняется анализ синтаксического соответствия кыргызского и русского языков, проводится сравнительный анализ грамматического строя языков, разрабатывается общая концептуальная схема функционирования автоматизированной системы перевода, а так же описываются принципы организации базы знаний системы с приведением её предлагаемой структуры с подробным описанием файлов базы и их форматов.

Общая концептуальная схема функционирования системы отображена на следующем рисунке **рис.1**

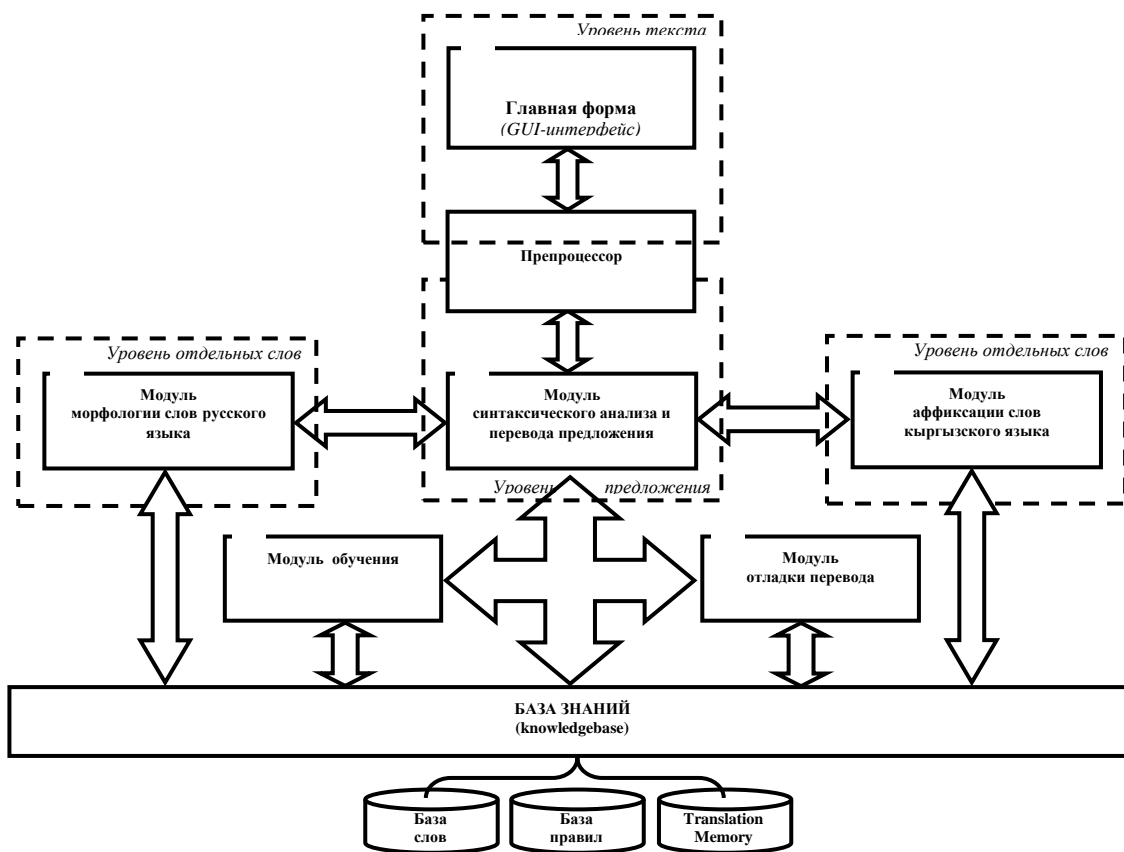


Рис. 1 Концептуальная схема системы

Основные элементы системы и их функциональные назначения:
главная форма-модуль – это модуль, имеющий графический интерфейс, который, по сути, связывает пользователя с ЭВМ. Главная процедура модуля работает на уровне текста – разбивает текст на отдельные предложения, и представляет их в виде двухмерной структуры (исходное предложение; переведенное предложение). Затем, предложения поочередно отправляются на предварительную обработку в препроцессор. После препроцессорной обработки предложения также поочередно отправляются на вход главной процедуры модуля;
препроцессор - модуль, содержащий процедуры для предварительной обработки входного текста;
модуль синтаксического анализа и перевода предложения – основной компонент системы перевода. Данный модуль, как видно из концептуальной

схемы, напрямую взаимодействует со всеми другими модулями системы. Сначала система выполняет поиск предложения, поступившего на вход, в базе Translation Memory (далее ТМ).

А затем, если его перевод не был найден, предложение представляется в виде структурированного графа, элементами которого являются слова и их различные морфологические и синтаксические свойства, а также, конечно, результат перевода. Элементы графа заполняются по мере получения данных от модулей морфологического анализа, а также результата синтаксического анализа предложения и, в итоге, результата перевода. В процессе перевода, модуль проверяет связь каждого анализируемого слова с соседними словами, используя базу ТМ. Он также в процессе перевода, может автоматически обращаться к системе обучения. После получения всех необходимых данных и формирования переведенного предложения, модуль передает результат перевода и, соответственно, управление главному модулю;

модуль морфологии слов русского языка – принимает на вход отдельные слова от синтаксического анализатора и проводит их морфемный анализ. Затем, на основе результатов морфемного анализа, модулем проводится морфологический анализ. Полученные результаты и управление передаются синтаксическому анализатору. Синтаксический анализатор записывает полученные данные в соответствующие элементы структурированного графа и, затем, отправляет на морфологический анализ следующее слово. Этот цикл продолжается до тех пор, пока все слова в предложении не будут проанализированы;

модуль морфологии слов кыргызского языка – принимает на вход отдельные слова кыргызского языка в начальной форме, непосредственно после их перевода. Процедуры данного модуля производят сопоставление морфологических свойств русского и кыргызского языков, затем выполняется процедура аффиксации. После этого, модифицированное слово (в нужной форме) и его свойства передаются модулю синтаксического анализа и перевода и записываются в соответствующие элементы графа;

модуль обучения – состоит из нескольких форм-модулей, включает в себя операции добавления новых слов в базу данных, новых пар схем предложения в базу правил, а также добавление пар предложений (оригинал-перевод) в базу ТМ. Все модули обучения вызываются модулем синтаксического анализа и перевода предложения автоматически, непосредственно в процессе перевода;

модуль отладки перевода – включает в себя систему оценки качества перевода, систему поиска альтернативных вариантов перевода, а также системы автоматизированной генерации синтаксических и морфологических правил;

база данных – является распределенной и представляет собой совокупность данных, связь между которыми установлена ассоциативная;

база правил – содержит в себе выраженные на формализованном языке правила построения слов и предложений;

база ТМ –память переводчика – также является распределенной. При анализе, система всегда сначала обращается к базам ТМ, и уже потом, при необходимости, к базам данных и правил.

Структура базы ТМ приведена на рис. 2.

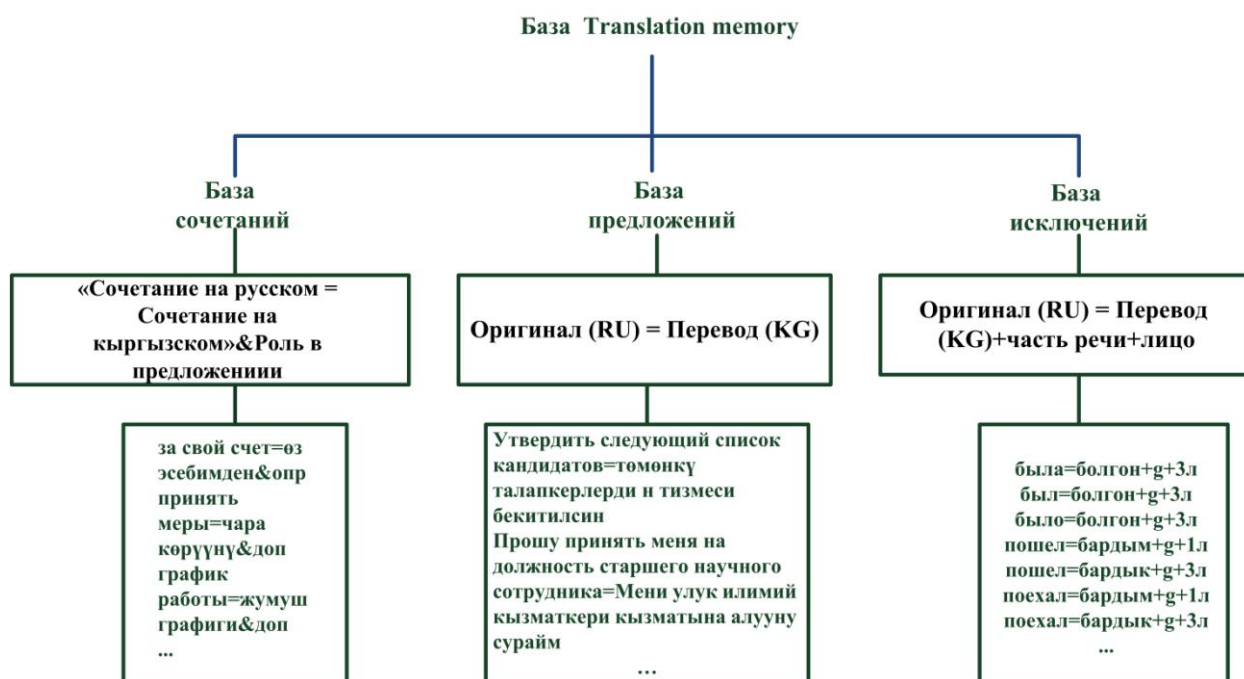


Рис.2. Структура базы Translation memory

Для организации процесса обучения и приобретения знаний система должна быть снабжена эффективно структурированной базой знаний. Исходя из таких требований к разрабатываемой системе, в ней реализована нижеописанная структура базы знаний (рис.3.).

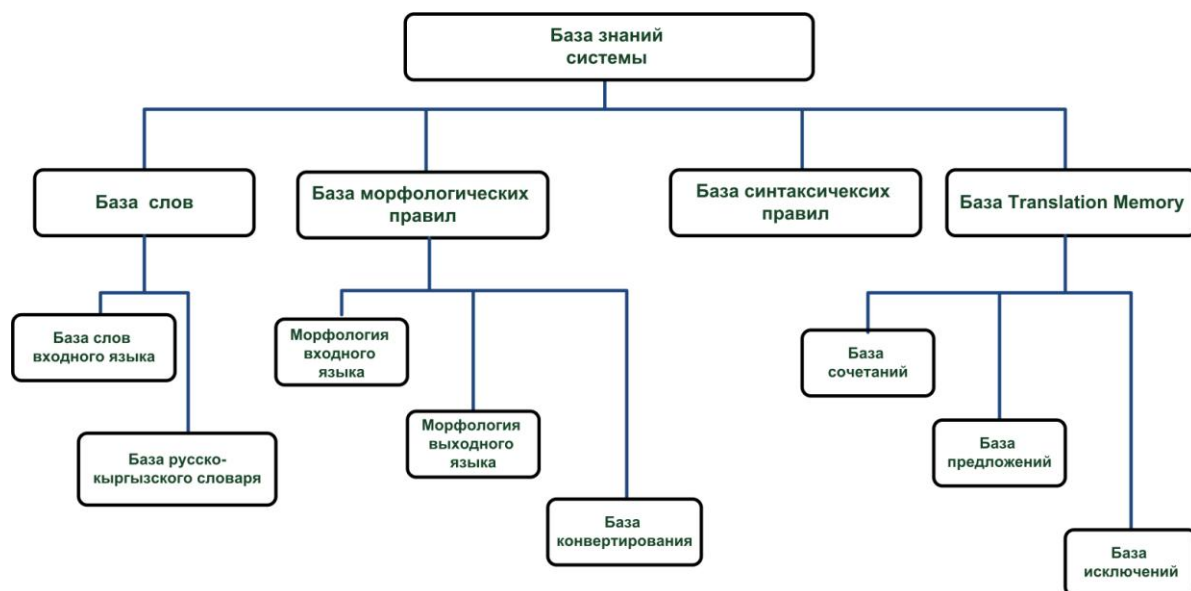


Рис.3. Структура базы знаний системы.

Где, база слов входного языка распределена по частям речи и включает в себя следующие составляющие:

- база существительных одушевленных;
- база существительных неодушевленных;
- база прилагательных;
- база наречий;
- база служебных частей речи;
- база имен собственных;
- база глаголов.

Данная база используется модулем морфологического анализа русских слов. Формат записи данных во всех перечисленных базах, кроме базы глаголов, совпадает – это слова русского языка в начальной форме,

расположенные по следующему принципу: одна строка – одно слово. База глаголов имеет следующий формат записи данных:

«Инфинитив = | словоформа(1)|словоформа(2)|словоформа(3)...»

Благодаря такому формату, при поиске начальной формы (инфинитива) глагола, системе достаточно найти словоформу анализируемого глагола в правой части правила и, тем самым, автоматически определить его инфинитив, обратившись к левой части этого же правила. Объяснить такую организацию можно тем, что определение начальной формы любой другой поддерживаемой части речи, система осуществляет благодаря основе слова, которая в большинстве случаев не изменяется, чего нельзя сказать о глаголах. Основа инфинитива глагола в большинстве его словоформ претерпевает изменение и, следовательно, системе становится невозможно по этой словоформе восстановить начальную форму глагола, если ее нет в базе.

Базы русско-кыргызского словаря распределены также по частям речи:

- существительные;
- прилагательные;
- наречия;
- местоимения;
- служебные части речи;
- глаголы.

Формат записи данных во всех базах, кроме глаголов следующий:

«Н.Ф. слова на русском = Н.Ф. слова на кыргызском».

Формат записи данных в базе глаголов:

«Инфинитив глагола (RU) = Глагол в форме 2л., повелительного наклонения(KG) & Глагол в форме 3л., повелительного наклонения(KG) & Глагол в инфинитиве (KG)».

В качестве варианта перевода, система выбирает один из трех содержащихся в базе вариантов перевода, в зависимости от того, в какой форме находится русский глагол, и какую роль он играет в предложении.

Базу морфологических правил можно условно разделить на следующие три типа баз:

- базу правил входной (русской) морфологии;
- базу правил выходной (кыргызской) морфологии;
- базу конвертирования русской морфологии в кыргызскую.

Структура базы морфологических правил имеет вид, отображенный на рисунке 4

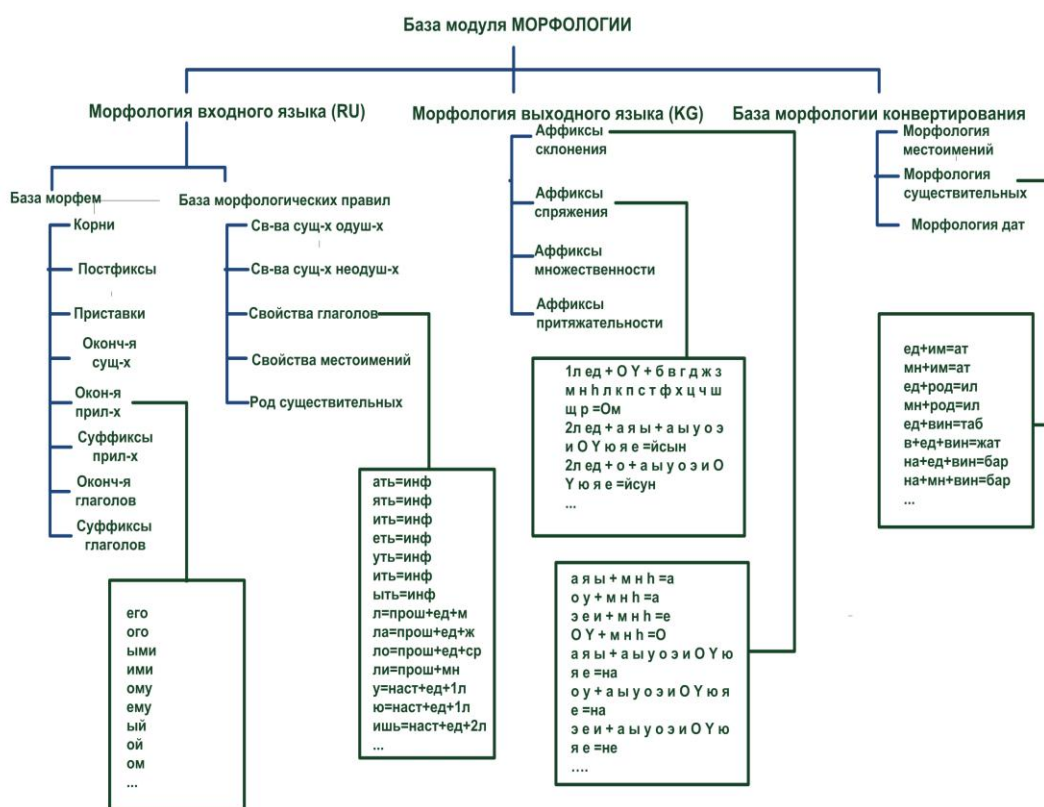


Рис.4. Структура базы модуля морфологии

Вышеперечисленные базы активно используются процедурой морфемного анализа слов, принцип работы которой подробно описан во второй главе.

В третьей главе описывается процесс разработки алгоритмов морфологического, синтаксического анализ, а так же алгоритма перевода предложений. Для полного отображения принципов функционирования

разработанной системы приводится схема функционирования системы, где детально отображается последовательность действий и принципы взаимодействия отдельных модулей системы с соответствующими компонентами базы знаний. Схема приведена ниже на рисунке 5.

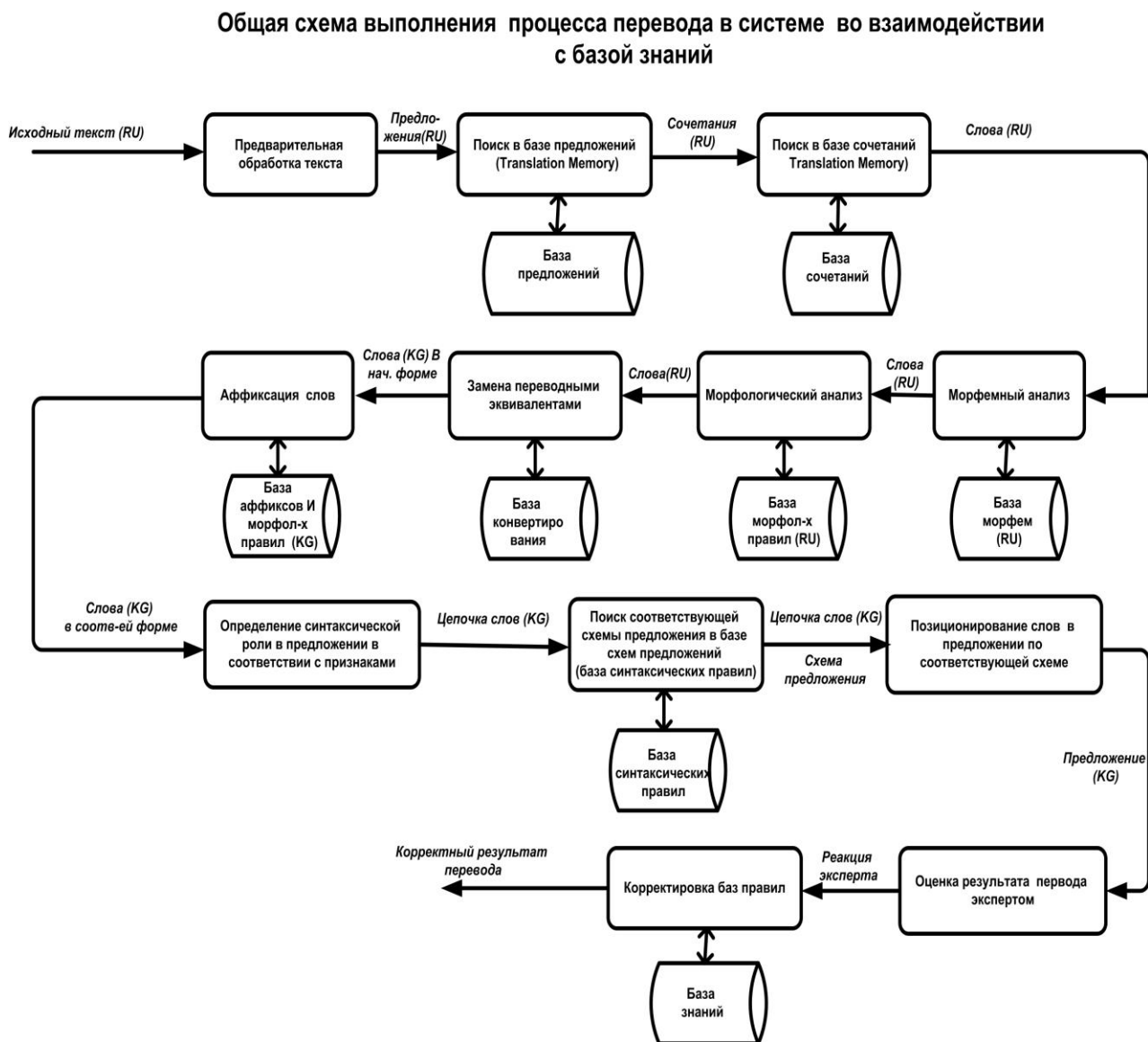


Рис 5. Общая схема выполнения процесса перевода.

Этап синтаксического анализа текста в процессе трансляции является одним из важных и сложных в реализации транслятора. Цель синтаксического анализа – определение роли отдельного слова в предложении и взаимосвязей между элементами предложения. Элементами предложения являются слова и словосочетания. Синтаксический анализатор транслятора первоначально определяет, каким членом предложения будет являться каждое из его слов (словосочетаний). Слова в предложении могут

являться главными и второстепенными членами предложения. Обозначения членов предложения аналогичны их обозначению в русском языке, но, тем не менее, имеют достаточно много различий, так как в русском языке члены предложения определяются по смыслу. Для описания синтаксических схем предложений в трансляторе приняты условные члены предложения (дополнения, определения и обстоятельства), которые определяются по соответствующим правилам. Данный механизм в трансляторе работает только для простых предложений (в которых может быть только одно подлежащее и одно сказуемое), но, с другой стороны абсолютно все сложноподчиненные и, особенно, сложносочиненные предложения можно без особого труда разбить на простые, а следовательно, можно сделать вывод, что теоретически, данный механизм является универсальным для обработки предложений любой сложности. Алгоритм синтаксического анализа представлен на рисунке 6.

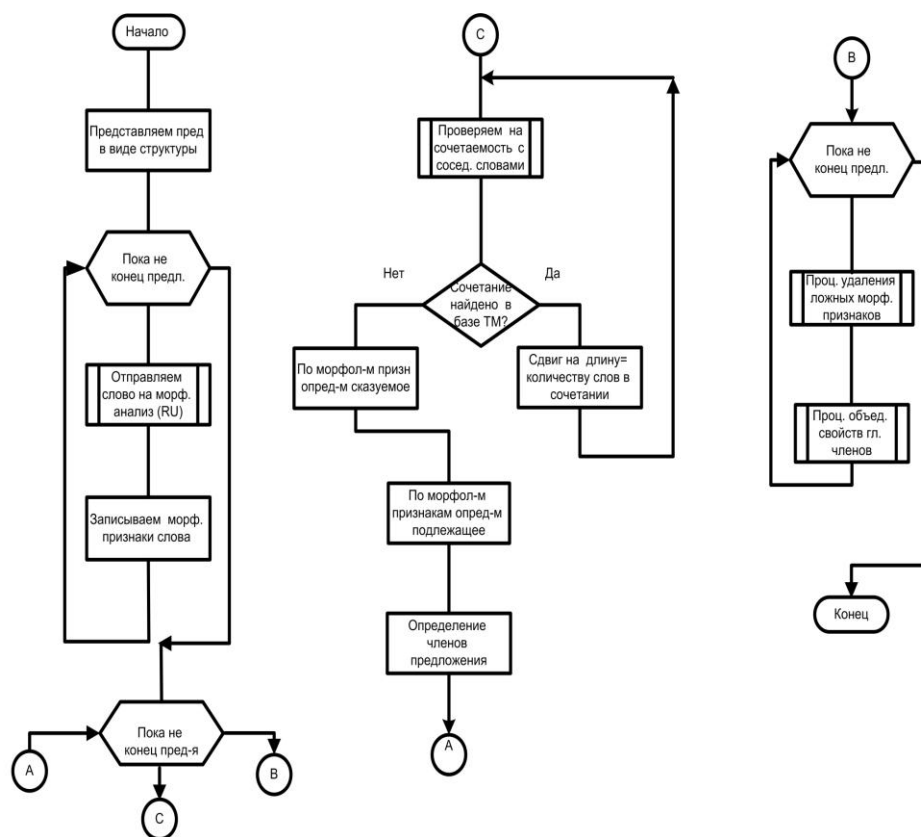


Рис. 6. Алгоритм синтаксического анализа предложений

После определения схемы предложения, система формирует схему предложения для выходного языка, а также объединяет свойства главных

членов, связывает предлоги с соответствующими им дополнениями, отбрасывает ложные морфологические свойства и проводит другую «связывающую» работу. После перевода всех слов, система расставляет переведенные слова в предложении, согласно выходной схеме. Алгоритм выполнения перевода предложения отображен на рисунке 7.

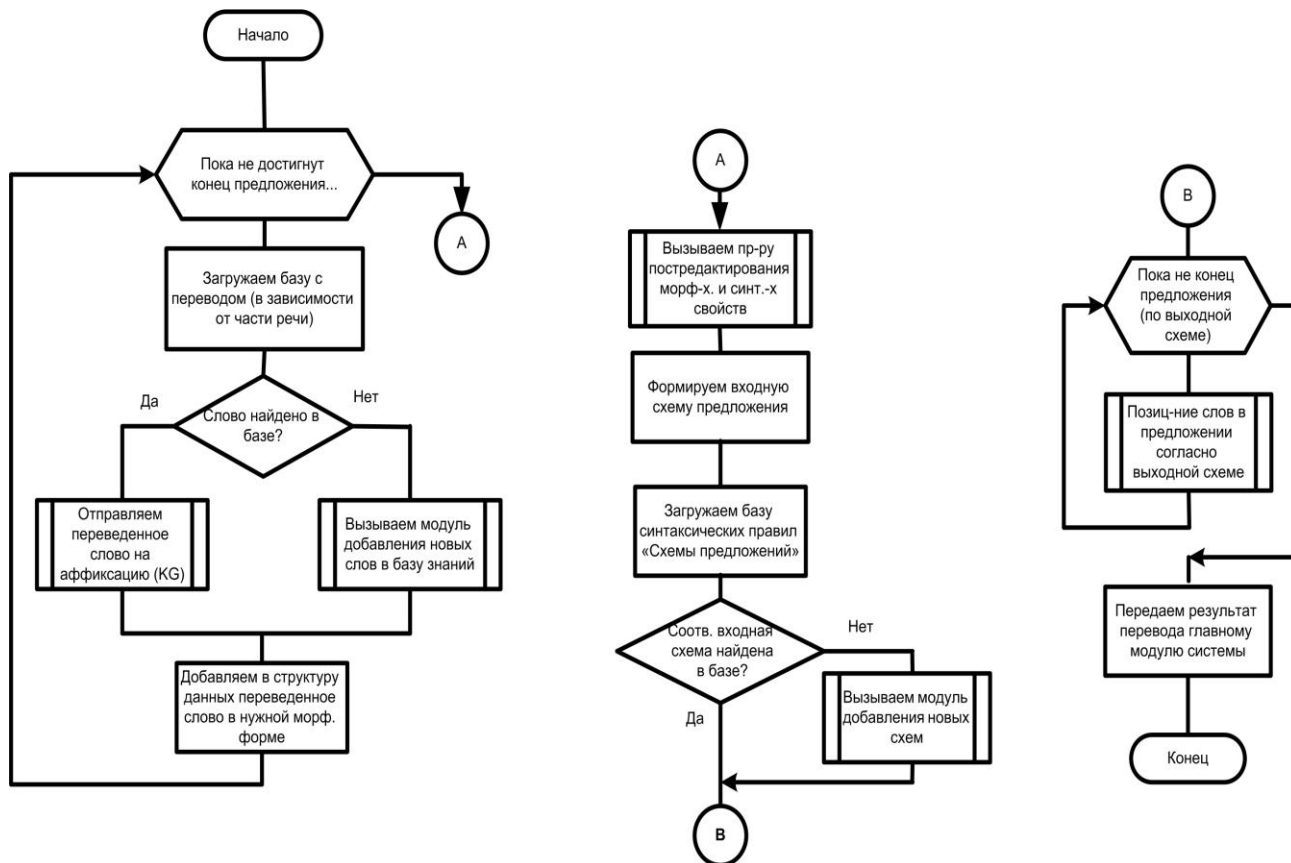


Рис.7. Алгоритм перевода предложения

В четвертой главе описан процессы разработки алгоритмов оценки качества перевода и отладки перевода, позволяющие реализовать механизмы приобретения знаний и обучения системы. Разработан дружественный интерфейс системы. Подробно описан процесс редактирования базы данных и правил.

В системе реализованы следующие процедуры, обеспечивающие непосредственно алгоритм обучения:

- а) добавление нового слова в базу данных;
- б) добавление нового синтаксического правила в базу правил;
- в) добавление нового морфологического правила в базу правил;

- г) добавление неделимого словосочетания в базу Translation Memory;
- д) добавление пары «Оригинал - Перевод» в базу Translation Memory;
- е) добавление словоформы, как исключение в базу Translation Memory.

Процедуры (а) и (б) могут вызываться как пользователем, так и, при необходимости, самой системой перевода.

К алгоритмам отладки перевода можно отнести:

- алгоритмы обучения системы;
- алгоритм оценки качества перевода.

Общий алгоритм процедуры оценки качества перевода приведен на рисунке 8.

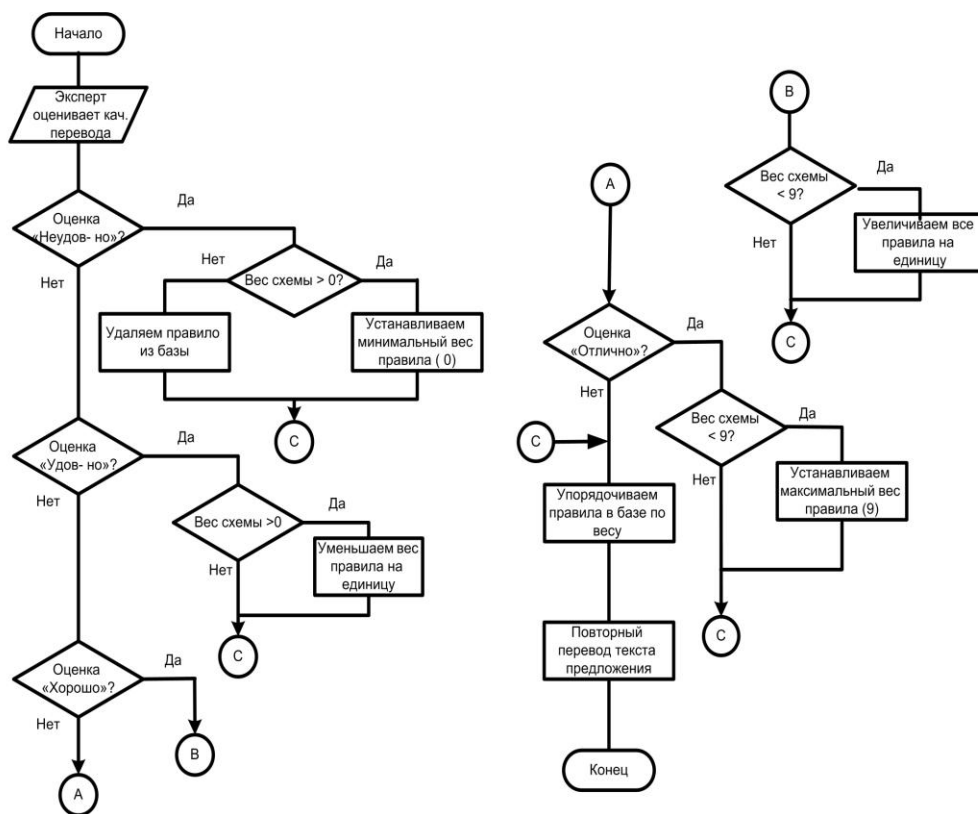


Рис. 7. Общий алгоритм процедуры оценки качества перевода.

В системе так же реализованы следующие процедуры, которые составляют процесс автоматизированного обучения:

- а) процедура генерации новых синтаксических правил;
- б) процедура генерации новых морфологических правил.

Процедура использует форму, имеющую удобный графический интерфейс, с помощью которой пользователь может сформировать новое

синтаксическое правило и, затем, добавить его в базу. В том случае, когда процедура вызывается автоматически, она определяет схему входного предложения, эксперту необходимо определить соответствующую выходную схему предложения, а также начальный вес сформированного синтаксического правила. После того, как пользователь введет все необходимые данные, система формирует новое правило и добавляет его в базу «Схемы предложений». После добавления нового синтаксического правила, текст переводится повторно, уже с учетом нового правила.

Сначала система повторно анализирует и переводит предложение, запоминая все отладочные данные. Если синтаксическое правило для данного типа предложений отсутствует в базе, система генерирует временную выходную схему, полностью идентичную входной. Если синтаксическое правило имеется в базе, система принимает за временную выходную схему этого правила. Затем, система формирует временный перевод предложения, расставляя в нем слова, согласно временно сгенерированной схеме. Эксперт определяет нужный порядок слов в выходном предложении, тем самым, создавая новую версию перевода предложения.

Система, используя процедуру пословного сравнения временного перевода и перевода пользователя, определяет выходную схему для предложения эксперту. Пользователь устанавливает относительный вес нового правила, после чего система добавляет новое правило в базу «Схемы предложений».

Процедура генерации новых морфологических правил способна генерировать правила для существительных, личных местоимений и их сочетаний с предлогами. Рассмотрим алгоритм работы данной процедуры на примере генерации правила для существительного.

Сначала, эксперт вводит существительное в любой форме (можно в сочетании с предлогом) на русском языке и сам же переводит его на кыргызский язык. Процедура отправляет русское слово на морфологический

анализ и, тем самым, определяет его морфологические свойства. Затем, процедура анализирует киргызское слово, сканируя базы аффиксов склонения. После того, как и русские и киргызские свойства определены, процедура генерирует морфологическую ассоциацию «рус.морф.св-ва = кирг.морф.св-ва», после чего, ее правильность оценивает эксперт. При отрицательной оценке эксперта, система вызывает процедуру обучения, благодаря которой эксперт может вручную составить новое правило, затем новое правило добавляется в базу. При положительной оценке эксперта, автоматически сгенерированное правило, также добавляется в соответствующую базу. Выше описанные алгоритмы генерации синтаксических и морфологических правил приводятся на рисунке 8.

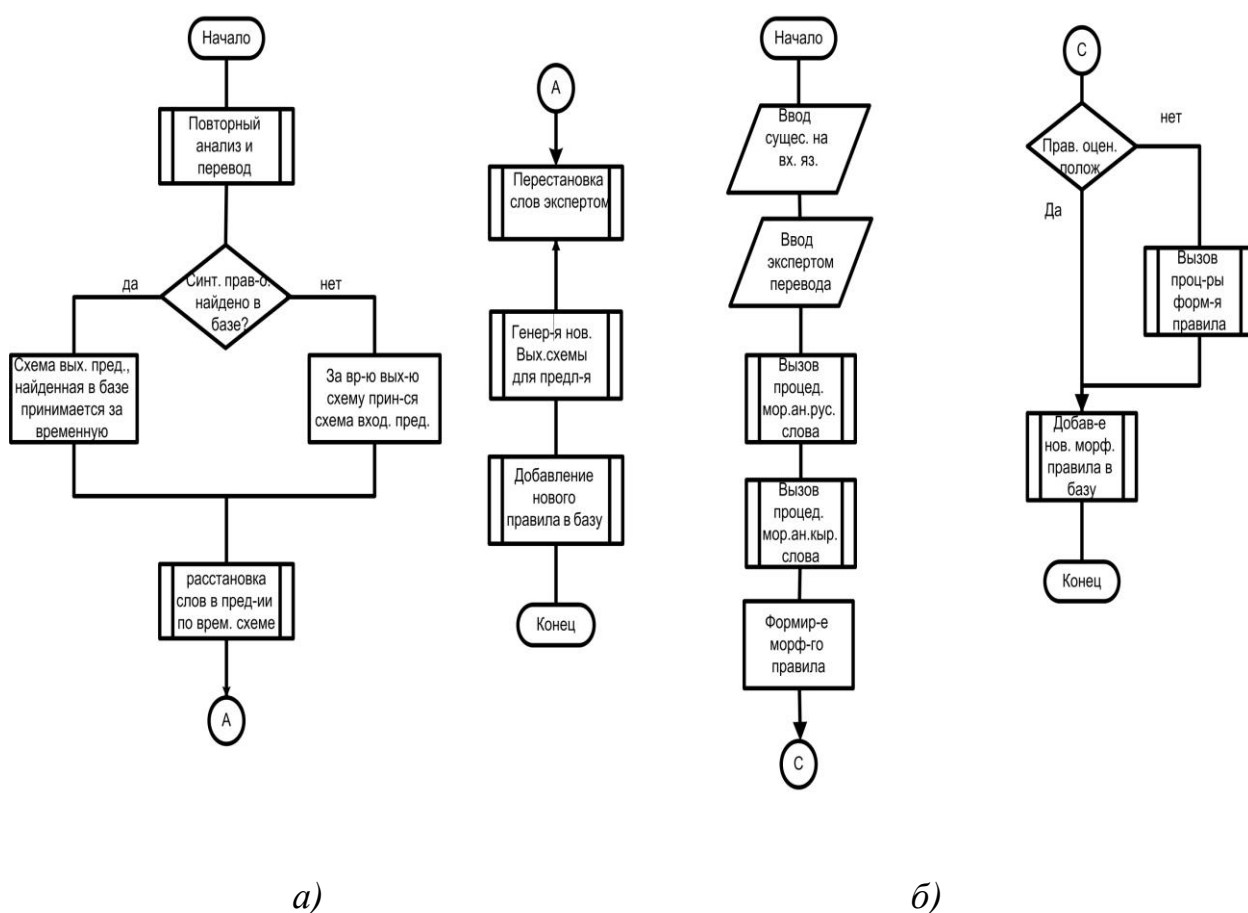


Рис. 8. Алгоритмы генерации новых правил (а-синтаксических; б-морфологических)

В заключении изложены основные научные и практические результаты, полученные в диссертационной работе.

В приложении приведены листинг разработанного программного кода, акт внедрения.

ВЫВОДЫ

В работе получены следующие результаты:

1. Разработан модуль морфологического анализа слов;
2. Разработан модуль синтаксического анализа и перевода предложения;
3. Разработан механизм обучения системы, который состоит из нескольких модулей, включающих в себя операции добавления новых слов в базу данных, новых пар схем предложения в базу правил, а также добавление пар предложений (оригинал-перевод) в базу Translation Memory .

4. Разработан и реализован алгоритм отладки процесса перевода, включающий в себя модуль оценки качества перевода, модуль поиска альтернативных вариантов перевода, а также модуль генерации синтаксических и морфологических правил.

5. Разработана автоматизированная система лингвистического перевода в русско-кыргызском направлении, способная накапливать знания и обучаться.

Список опубликованных работ:

1. **А.с. 268** Кыргызпатент. Автоматизированная система русско-кыргызского перевода текста [Текст]/Н.А.Исраилова // №268, 25.04.2013;заявл.03.04.2013, опубл. 05.2013, Бюл.№5-17с.
2. **Исраилова Н.А.** Организация синтаксического анализа в трансляторах [Текст] /Н.А.Исраилова //Вестник Восточно-Казахстанского Государственного Технического Университета им. Д Серикбаева – Усть-Каменогорск– №1,март 2012.—С.101-104.
3. **Исраилова Н.А.** Организация морфологического анализа в трансляторах [Текст] /Н.А.Исраилова //Вестник Восточно-Казахстанского Государственного Технического Университета им. Д Серикбаева –Усть-Каменогорск– №1,март 2012.— С.97-101.
4. **Исраилова Н.А.** Building a bridge between IT and not-IT specialists-TEMPUS ERAMIS project. [Текст]/Исраилова Н.А., Марек Милош// Известия КГТУ им. И.Раззакова-2011.-№24- С.233-236.
5. **Исраилова Н.А.** Алгоритмы отладки процесса трансляции [Текст] /Н.А.Исраилова // Известия КГТУ им. И.Раззакова—2011.- №22-С.278-279.
6. **Исраилова Н.А.** Алгоритмы отладки процесса перевода [Текст] /Н.А.Исраилова, М.А. Шабданов // Материалы 53-НТК молодых ученых, посвященный 100-летию И. Раззакова— 2011.-С.104-107.
7. **Исраилова Н.А.** Реализация синтаксического анализа в трансляторах. [Текст] /Н.А.Исраилова, М.А. Шабданов // Материалы 52-НТК молодых ученых, посвященный 100-летию И. Раззакова— 2010.—С. 214-217.
8. **Исраилова Н.А.**Создание оболочки экспертной системы. [Текст] /Н.Т.Иманакунуова, Н.А.Исраилова //Материалы 52-НТК молодых ученых, посвященный 100-летию И. Раззакова— 2010.—С.224-227.
9. **Исраилова Н.А.** Принципы организации морфологического анализатора в трансляторе[Текст] /Н.А.Исраилова // Известия КГТУ им. И.Раззакова-2010.-№20- С.233-236.
- 10.**А.с. 209** Кыргызпатент. Экспертная система языкового перевода [Текст]/Н.А.Исраилова // №209, 21.05.2010;заявл.07.04.2010, опубл. 07.2010, Бюл.№7-25с.

11. **Исраилова Н.А.** Алгоритмы обучения модели искусственной нейронной сети [Текст] /Н.А.Исраилова // Известия КГТУ им. И.Раззакова – 2006.-№10—С.262-264.
12. **Исраилова Н.А.** Анализ инструментальных средств имитационного моделирования[Текст] /Н.А.Исраилова // Материалы международной конференции, посвященной 20-летию кафедры ЭВМ- ФПИ Бишкек,2003.—С.210-213.
13. **Исраилова Н.А.** Оболочка имитационного моделирования [Текст] /Н.А.Исраилова // Материалы международной конференции «ИНФООПТИКА-2002», Бишкек, 2002.— С.150-153.

Исраилова Нелла Амантаевнанын

05.13.06-«Технологиялык процесстерди жана өндүрүштү автоматташтыруу жана башкаруу» адистиги боюнча «Лингвисттик которуунун автоматташтырылган системасын иштеп чыгуу» аттуу темасында аткарылган диссертациясынын

КЫСКАЧА МАЗМУНУ

Негизги сөздөр: автоматташтырылган которуу, машиналык которуу, эксперттик система, морфологиялык талдоо, синтаксистик талдоо, билим базасы, которгуч, автоматташтырылган система.

Изилдөө объектиси: лингвистикалык которуу процессин автоматташтыруу технологиясы

Иштин максаты: лингвисттик которуунун автоматташтырылган системасын иштеп чыгуу.

Изилдөө ыкмалары: Алдыда коюлган маселени чечүүдө иште талдоо ыкмалары, лингвистикалык мыйзамченемдүүлүктөрдү туюндуруучу математикалык үлгүлөрдү түзүү үчүн моделдөө элементтери колдонулган. Жана дагы объектиге багытталган программалоо ыкмалары колдонулган.

Иштин негизги натыйжалары: билим топтоо жана аны колдонуу режимдеринде иштей ала турган, котормочунун түзүмүн уюштуруунун жаңы принцибине негизделген, орусчадан-кыргызчага багытындагы тексти которуунун автоматташтырылган системасынын иштелип чыгышы.

Изилдөөнүн натыйжаларын колдонуу: иштелип чыккан лингвисттик которуунун автоматташтырылган системасы И. Раззаков атындагы КМТУнун окуу процессинде, КР ,кмөтүнүн аппаратынын регламент жана иш кагаздарын жүгүртүү бөлүмүндө колдонууга киргизилди.

Колдонуу тармагы: изилдөөнүн натыйжалары жана иштелип чыккан система расмий иш кагаздарын орус тилинен кыргыз тилине которууну аткаруучу ведомстволордо колдонулушу мүмкүн.

РЕЗЮМЕ

диссертации Исраиловой Неллы Амантаевны на тему: "Разработка автоматизированной системы лингвистического перевода" на соискание ученой степени кандидата технических наук по специальности 05.13.06 - "Автоматизация и управление технологическими процессами и производствами"

Ключевые слова: автоматизированный перевод, машинный перевод, экспертная система, морфологический анализ, синтаксический анализ, база знаний, транслятор, автоматизированная система.

Объект исследования: технология автоматизации процесса лингвистического перевода.

Цель исследования: разработка автоматизированной системы лингвистического перевода.

Методы исследования: при решении поставленных задач в работе использованы аналитические методы. Применены элементы моделирования для построения математических моделей, описывающих лингвистические закономерности, а также методы объектно-ориентированного программирования.

Основные результаты работы и их новизна: разработана автоматизированная система перевода текста в русско-кыргызском направлении, реализованная на базе нового принципа организации структуры переводчика.

Использование результатов исследования: автоматизированная система лингвистического перевода внедрена в работу отдела регламента и делопроизводства аппарата Правительства КР и в учебный процесс КГТУ им. И. Раззакова.

Область применения: Результаты исследования и разработанная система могут быть использованы в работе ведомств, занимающихся переводом официальной документации с русского языка на кыргызский.

SUMMARY

dissertation of Israilova Nella on the theme: "Development automated system of linguistic translation" for the degree of candidate of technical sciences, specialty 05.13.06 - "Automation and management of technological processes and productions"

Keywords: automated translation, machine translation, expert system, morphological analysis, syntactic analysis, knowledge base, translator.

Object of the research: automate the process of linguistic technology translate.

Purpose of the research: development of an automated linguistic translation development of an automated linguistic translation that can translate text from Russian into Kyrgyz.

Research methods: in solving the problems in the analytical methods used. There are elements of modeling for the construction of mathematical models describing the linguistic patterns. As well as methods of object-oriented programming.

The main results are: an automated system for translating text in Russian-Kyrgyz direction, implemented on the basis of the principle of a new organization structure of the interpreter, which allows it to operate in two modes: the acquisition of knowledge and in the mode of their application.

Using the results of the study: the introduction of manufactured developed automated language translation of the educational process KSTU. I.Razzakov, introduced the work of the translation department of the Government of the Kyrgyz Republic.

Scope: the results of the research and the developed system can be used in the structures involved in the translation of texts official style.