

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

Шихиев
Фуад Шукурович

ФОРМАЛИЗАЦИЯ И СЕТЕВАЯ ФОРМУЛИРОВКА
ЗАДАЧИ СИНТАКСИЧЕСКОГО АНАЛИЗА.
(Модели и алгоритмы)

05.13.11. – математическое и программное обеспечение вычислительных машин,
комплексов и компьютерных сетей

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Санкт-Петербург
2006

Работа выполнена на кафедре информатики и вычислительной техники математического факультета Дагестанского государственного университета.

Научный руководитель: кандидат технических наук,
доцент

Шамаев Алексей Эдуартович

Официальные оппоненты: доктор физико-математических наук,
профессор Романовский Иосиф Владимирович
кандидат физико-математических наук,
доцент Кривцов Александр Николаевич

Ведущая организация: Санкт-Петербургский
экономико-математический институт РАН

Защита диссертации состоится “ ___ ” _____ 2006 года в ___ часов на заседании диссертационного совета Д212.232.51 по защите диссертаций на соискание ученой степени доктора наук при Санкт-Петербургском государственном университете по адресу: 198504, Санкт-Петербург, Старый Петергоф, Университетский пр., д. 28, математико-механический факультет Санкт-Петербургского государственного университета.

С диссертацией можно ознакомиться в Научной библиотеке Санкт-Петербургского государственного университета по адресу: 199034, Санкт-Петербург, Университетская наб., д. 7/9.

Автореферат разослан “ ___ ” _____ 2006 года.

Ученый секретарь
диссертационного совета
доктор физико-математических наук,
профессор

Б.К.Мартыненко

Общая характеристика работы

Актуальность темы

Работа компьютера строго предопределена алгоритмом и данными, которые предварительно записываются в его память человеком. Правила и приемы, определяющие способность человека к хранению и обработке информации, не всегда удается понять и расписать. К числу таких способностей человека относится его *языковая способность*, которая, в частности, выполняет коммуникативные функции. «Недопонимание» между человеком и компьютером при общении появляется тогда, когда каждый из них измеряет информацию «по своей мерке».

Выявление формальных структур в естественном языке (ЕЯ), формализация языка в целом, построение конструктивной теории и компьютерной модели языка являются актуальными направлениями компьютерной лингвистики, ибо работы, проводимые в этом направлении, ориентированы на увеличение числа пользователей компьютера и расширение круга задач, решаемых на компьютере.

Цель работы

Под *конструктивной теорией* языка понимается теория, каждое понятие которой имеет *конструктивное определение*, а методы преобразования понятий задаются *алгоритмом*. *Конструктивно-определенное понятие* представляет собой структуру, для которой известны алгоритмы ее построения (синтеза) и распознавания (анализа). Конструктивная теория имеет *компьютерную модель*, если *понятия-структуры* этой теории представимы в памяти компьютера. Задачей диссертационного исследования является построение *компьютерной модели ЕЯ*, более полное изучение задачи синтаксического анализа (СА) предложений ЕЯ и подготовка программы для синтаксического анализа предложений русского языка.

Основные результаты

Автором разработана программа (в среде Delphi), в которой реализована формальная модель синтаксиса русского языка и алгоритмы, поддерживающие морфологический анализ словоформ и синтаксический анализ. Результатом синтаксического анализа последовательности словоформ (предложения-последовательности) является словосочетание (предложение-дерево), которое в виде корневого дерева отображается на форме проекта.

Научная новизна

Новизна исследования заключается в следующем. (В этом пункте курсивом выделены термины, которые будут определены ниже)

1. Показана возможность конструктивного моделирования синтаксиса ЕЯ.
2. Приводится конструктивное определение ЕЯ. Введенное в диссертационной работе понятие *синтаксической формы* позволяет строить *отображение* элементов синтаксиса во множество элементов семантики, которое, по определению, представляет собой *язык*.
3. Благодаря простому и компактному представлению *синтаксических функций и словосочетаний, парадигм и синтаксического словаря, синтаксических отношений и словосочетаний*, которые связаны в единую и взаимосвязанную структуру, достигается гибкость и открытость исследуемой модели синтаксиса.
4. Синтаксис задается ориентированным графом, множеством вершин которого являются слова и словоформы языка, а каждая дуга представляет собой связанную пару словоформ.
5. Основные понятия синтаксиса определяются через элементы графа, что позволяет применить к задаче синтаксического анализа богатый арсенал сетевых методов.
6. Алгоритмы синтеза и анализа элементов синтаксиса ЕЯ реализованы в среде Delphi и апробированы для анализа словоформ и словосочетаний русского языка.

Примечание [БМ1]: Разве это не было известно раньше? Интересно знать, каким образом Вы предлагаете:
- описывать множество всех предложений, составляющих язык?
- представлять структуру любого предложения языка?

Практическая и теоретическая ценность

Теоретическая значимость работы заключается в том, что в диссертации описана конструктивная теория ЕЯ. В частности, подробно изучены свойства синтаксических функций и сформулированы требования к синтаксическим категориям, при которых слова и словоформы языка могут быть сохранены в словаре в простом и содержательном формате. В рамках рассматриваемой модели синтаксиса можно описать синтаксис любого **письменного** языка.

Примечание [БМ2]: Может ли эта модель применяться к нероглифической письменности?

Практическая ценность работы состоит в том, что алгоритмы решения задач анализа и синтеза словоформ и словосочетаний не зависят от объема словаря и содержания синтаксических отношений (принцип открытости модели). Следовательно, программа, подготовленная для анализа предложений русского языка, может быть использована для синтаксического анализа предложений любого языка, в котором элементы парадигм отличаются только постфиксами. (Словоформы связаны отношениями (1) из содержания работы в главе первой.) Исходные данные программы-синтаксиса (лексика и синтаксические отношения) могут быть изменены программно, следовательно, имеется возможность для экспериментального исследования и сравнительного анализа различных моделей синтаксиса. Синтаксический словарь, подготовленный в ходе работы с *программой-синтаксисом*, подобен известному словарю А.А. Зализняка и может быть использован в текстовых редакторах для проверки орфографии слов и словоформ. Классификация предложений текста (художественного произведения) по их *синтаксическим формам* показывает частоту использования автором произведения той или иной синтаксической формы и характеризует его индивидуальность.

Апробация работы

Полученные результаты печатались в изданиях Дагестанского научного центра РАН и докладывались на конференциях, проводимых ДНЦ РАН, ДГУ и ДГПУ. Преподавателями Дагестанского госуниверситета при участии

соискателя подготовлена к печати монография «Формальная теория ЕЯ». Материалы диссертации используются в специальных курсах по моделированию синтаксиса и семантики, в курсовых и дипломных работах. Процедуры морфологического анализа словоформ успешно используются на занятиях по информатике со студентами-филологами.

Публикации

Основные результаты представлены в работах [1]-[6].

Структура и объем диссертации

Диссертация объемом 171 машинописных страниц, содержит введение, пять глав и заключение, список литературы (133 наименований), 30 таблиц, 15 рисунков.

Содержание работы

Во введении обоснована важность и актуальность темы диссертации, сформулированы цели работы и основные задачи, которые необходимо решать для ее достижения, характеризуется научная новизна и практическая ценность работы, кратко излагаются основные результаты.

В первой главе диссертации приводится описание синтаксиса как формальной системы. Синтаксис ЕЯ представляет собой систему, которая задается множеством исходных элементов и набором операций, определенных над данным множеством. *Исходными элементами* ЕЯ являются слова и словоформы. *Унарные операции* переводят слова в словоформы. Каждая синтаксическая категория ЕЯ представляет собой вектор-функцию, элементами которой являются *синтаксические функции* (унарные операции) с общей областью определения. Например, категория числа $H = (h_1, h_2)$ определена на слове $x = \text{синий}$ и имеет значения $H(\text{синий}) = (h_1(\text{синий}), h_2(\text{синий})) = (\text{синий}, \text{синие})$. Категории рода $R = (r_1, r_2, r_3)$ и падежа $P = (p_1, p_2, p_3, p_4, p_5, p_6)$

определены на слове x и $R(\text{синий}) = (r_1(\text{синий}), r_2(\text{синий}), r_3(\text{синий})) = (\text{синий}, \text{синее}, \text{синяя})$, а $P(\text{синяя}) = (\text{синяя}, \text{синей}, \text{синей}, \text{синюю}, \text{синей}, \text{синей})$.

Бинарные отношения связывают пару словоформ, например, слова **дому** и **сладкому** связаны синтаксическим отношением, которое называется *согласованным определением*. Отношения между словоформами названо *синтаксическими* потому, что они определяются синтаксической формой членов данного отношения. В грамматике каждого ЕЯ известен конечный набор допустимых бинарных синтаксических отношений R . Если множество слов и словоформ языка обозначается через S^* , то нетрудно заметить, что синтаксические отношения задают граф (S^*, R) . Если множество S^* делить на подмножества (синтаксические группы), а на каждое синтаксическое отношение накладывать дополнительное условие, требующее принадлежность членов отношения к определенным синтаксическим группам, то появится возможность регулировать (сужать или расширять), группировать и нумеровать синтаксические отношения.

Если за конструктивную модель синтаксиса взять ориентированный граф (S^*, R) , а всякое корневое дерево в синтаксисе (S^*, R) называть *словосочетанием*, а словосочетание, корнем которого является слово, обозначающее действие, называть *предложением*, то можно было бы считать процесс построения модели синтаксиса завершенным и приступить к ее реализации на компьютере. Однако при таком упрощенном и интуитивном определении модели появятся проблемы, связанные с обоснованием сделанных построений и реализацией задачи синтаксического анализа. Поэтому вводятся следующие формальные определения элементов синтаксиса.

а. Каждая категория $f = f[1..n] = (f_1, f_2, \dots, f_n)$ имеет *область определения* D_f . Слово x принадлежит D_f , если в последовательности $f_1(x), f_2(x), \dots, f_n(x)$ имеются хотя бы **два** различных элемента. Элементы $f(x)$ называются *словоформами* или *формами* слова x по категории f . Если на слове x определены категории f, g, \dots, h , то у векторов $f(x), g(x), \dots, h(x)$ есть **общий элемент**, который называется *исходной формой* слова x . Исходные формы слов из S^* образуют множество S ,

Примечание [БМ3]: Слово "сани":
ЧИСЛО (сани) = (сани, сани);
РОД (сани) = (сани, сани, сани);
ПАДЕЖ (сани) = (сани, ..., саниями, санях, ...);
Так что, ЧИСЛО или РОД не имеют области определения?

Примечание [БМ4]: Эти векторы из разных пространств. Их координаты – значения из S^* . Надо понимать вектор как упорядоченное конечное множество, в котором могут встречаться одинаковые элементы.

которое называется *лексикой* языка. Категория f называется *словоизменяющей* для слов из D_f .

Если $f = (f_1, f_2, \dots, f_n)$ и $h = (h_1, h_2, \dots, h_m)$ – различные категории, то $f_i(x) * f_j(x) = f_i(x)$ и $f_i(x) * h_j(x) = h_j(x) * f_i(x)$ при всех $i = 1..n, j = 1..m$. Пусть на слове x определены категории $f(x), g(x), \dots, h(x), r(x)$, каждая из категорий, начиная со второго, определена на области значений предыдущей категории. Тогда суперпозиция функций $f_i(x) * g_j(x) * \dots * h_k(x) * r_q(x)$ однозначно определяет форму слова x . Нетрудно заметить, что $t = f_i * g_j * \dots * h_k * r$ есть категория. Категория вида t при всевозможных значениях индексов i, j, \dots, k называется *парадигмой-категорией*. Множество значений парадигмы-категории $t(x)$ образует *парадигму* слова x по t . Элемент $t_q(x)$ парадигмы $t(x)$ обозначается через (x, i, j, \dots, k, q) , и говорят, что словоформа $t_q(x)$ имеет *векторную запись* (x, i, j, \dots, k, q) или (i, j, \dots, k, q) . Если X – некоторое множество слов, то через (X, i, j, \dots, k, q) обозначается множество форм (x, i, j, \dots, k, q) при всех x из X . Можно предположить, все элементы парадигмы попарно различны.

б. Морфологическое правило, по которому строятся элементы x_1, x_2, \dots, x_n данной парадигмы, таково:

$$x_1 = x_0 + s_1, x_2 = x_0 + s_2, \dots, x_n = x_0 + s_n, \quad (1)$$

где x_0 – наибольшая общая, левая вырезка (неизменяемая, левая часть) словоформ x_1, x_2, \dots, x_n , а s_1, s_2, \dots, s_n – постфиксы, представляющие собой правые вырезки из словоформ x_1, x_2, \dots, x_n , удовлетворяющие (1). Вектора постфиксов (s_1, s_2, \dots, s_n) для всех слов лексики и для всех парадигм хранятся в специальной *таблице постфиксов* ТРФ. Для русского языка таблица ТРФ содержит около тысячи элементов.

Исходная форма x_1 слова входит в каждую парадигму данного слова в качестве первого элемента. Если в словаре хранить исходную форму x_1 слова, длину неизменяемой части x_0 и номер k_0 строки таблицы ТРФ, в которой хранится вектор (s_1, s_2, \dots, s_n) , то по этим трем параметрам легко восстанавливаются словоформы x_1, x_2, \dots, x_n . Таким образом, словарная статья *электронного синтаксического словаря* содержит само слово в исходной форме

x_1 и последовательность пар $(|x_0|, k_0), (|x_1|, k_1), \dots, (|x_m|, k_m)$, каждая из которых вместе с x_1 определяет элементы одной парадигмы.

с. Предполагается, что синтаксические функции, парадигмы-категории и элементы парадигм упорядочены и пронумерованы. Две словоформы x и y связаны *бинарным отношением* (i, j, \dots, k, q) с номером n , которое обозначается выражением $x(y^n)$, если *зависимый член* этого отношения y имеет векторную запись n . То, что слово y имеет форму n , уже перечисляет его к какой-то группе слов. В определении связанной пары $x(y^n)$ указываются те множества слов X и Y , к которым принадлежат *главный член* x и *зависимый член* y , соответственно. Такое определение связанной пары дает возможность регулировать само бинарное отношение $X * Y^q = X * (Y, i, j, \dots, k, q)$, которое далее называется *синтаксическим отношением* и обозначается тройкой (X, Y, n) . На форму главного члена не накладываются никакие ограничения. Если $\text{num}(X)$ и $\text{num}(Y)$ – номера множеств X и Y , то целочисленный вектор $(\text{num}(X), \text{num}(Y), n)$ или $(\text{num}(X), \text{num}(Y), i, j, \dots, k, q)$ задает синтаксическое отношение. *Таблица синтаксических отношений* (ТСО) состоит из векторов, задающих синтаксические отношения. ТСО задает ориентированный *граф синтаксических отношений* (ГСиО). Под синтаксисом понимает ГСиО $G = (S^*, R)$

Пусть X и Y – имена существительные, $Z = \{\text{хотеть, думать}\}$, $U = \{\text{писать, бежать, строить}\}$, $a = (1, 2)$ – векторная форма имени существительного в единственном числе и родительном падеже, $b = (1, 1, 3)$ – векторная форма глагола в прошедшем времени, в единственном числе, в третьем лице, $c = (0, 0, 0)$ – векторная запись исходной формы слова. Тогда $(X, Y, a) = \{(\text{дом, моды}), (\text{тетрадь, Маши}), \dots\}$, $(X, U, b) = \{(\text{ученик, пишет}), (\text{заяц, бежит}), (\text{ласточка, строит}), \dots\}$, $(Z, U, c) = \{(\text{хотел, писать}), (\text{хотеть, строить}), (\text{думали, бежать}), \dots\}$.

d. Корневое дерево $T = (V, Q)$ в графе G называется *словосочетанием-деревом* в синтаксисе (S^*, R) . Последовательность из вершин словосочетания-дерева, в которой его вершины расположены в том порядке, в котором они встречаются при обходе этого дерева, образуют *словосочетание-*

последовательность. Если вершина при обходе дерева встречается более одного раза, то все остальные ее вхождения, кроме первого, удаляются. Задача построения словосочетания-дерева по словосочетанию-последовательности называется задачей *синтаксического анализа последовательности словоформ* (САПС).

Во второй главе диссертации исследуются особенности синтаксиса предложений русского языка. Предложение есть словосочетание, корнем которого могут быть только слова из заданного множества. Корнем предложения русского языка может быть любое слово из перечисленных ниже групп: *глагол в личной форме, глагол в повелительном наклонении, причастие в краткой форме, имя прилагательное в краткой форме и в сравнительной степени*.

Корень предложения называется *сказуемым*. *Подлежащим* называется член предложения, который является именем и связан со сказуемым как зависимый член в форме именительного падежа. На различных примерах показано, что приведенное в диссертации определение предложения не сужает множество *простых предложений* русского языка, несмотря на различия в определении главных членов предложения. Два члена предложения, связанные с одним и тем же словом по одному и тому же синтаксическому отношению, являются *однородными членами* предложения. В этой главе показано, каким образом можно устранить нарушение древовидности предложения, которое может быть нарушено при наличии однородных членов в предложении. Присутствие сказуемого (корня предложения-дерева) в предложении является обязательным условием. Предусмотрены правила восстановления сказуемого в случае его отсутствия.

Структура предложения зависит от содержания ТСО. В диссертации приводятся точные определения и других членов предложения. Разнообразие синтаксических отношений зависит от состава лексики. Для изучения и демонстрации возможности формального синтаксиса, необходимо и достаточно включить в словарь следующие классы слов: 1. глаголы и именные части речи, 2. местоимения, предлоги, наречия, союзы, частицы и связки.

Важным этапом синтаксического анализа последовательности словоформ является *сегментация* членов последовательности. Рассматриваются связи между *сегментами* и *фрагментами* в предложениях русского языка.

В третьей главе рассматриваются различные методы решения задачи САПС, которая сводится к построению в ГСиО корневого дерева T с заданным множеством вершин V . В общем виде эта задача решается *методом перебора* и является NP-полной задачей, но, используя особенности ГСиО, можно найти достаточно эффективные алгоритмы ее решения. Чтобы задача САПС была всегда разрешимой, возможны ее модификации. Например: в ГСиО найти корневое дерево, вершинами которого являются наибольшее число первых элементов заданной последовательности словоформ.

Подавляющее большинство слов русского языка имеет десятки словоформ. Поэтому ГСиО может иметь более миллиона вершин и десятки миллионов дуг. Используя результаты главы 1, ГСиО представлен в неявной форме, практически, не занимает место в памяти компьютера. На «восстановление» словоформ (вершин графа) путем *морфологического анализа слов* и на просмотр ТСО тратится определенное время, но оно измеряется миллисекундами и почти незаметно.

Приведенные в этой главе оценки сложности задачи СА при полном переборе возможных вариантов допустимых решений показывают отсутствие «комбинаторного взрыва» в алгоритмах СА последовательности из полсотни словоформ. Для решения задачи САПС в диссертации используются сетевые методы. Для отсева заведомо неперспективных вариантов наращивания строящегося предложения-дерева используются различные свойства задачи, в частности, *фрагментность сегментов*, согласно которого каждый *сегмент* (вершины ветви) является *фрагментом* предложения-последовательности.

Более подробно рассмотрены три алгоритма построения дерева в ГСиО. Метод «сборки ветвями» искомого дерева напоминает известный метод пузырька. Если в рассматриваемой последовательности много неизменяемых частей речи, то имеет смысл строить дерево «снизу вверх», с висячих вершин до

корня. Метод «сверху вниз» эффективнее при поиске одного из возможных решений задачи СА. Условие древовидности словосочетания позволяет не только формализовать понятия и конструкции синтаксиса, благодаря древовидности предложения алгоритмы СА обретают почти линейную сложность.

В четвертой главе приводится описание программы, которая осуществляет синтаксический анализ словосочетаний и предложений русского языка. Отдельно рассмотрены алгоритмы анализа и синтеза словоформ и словосочетаний. Несмотря на очевидную простоту алгоритма морфологического анализа словоформ, при реализации его возникает ряд проблем, связанных с представлением информации в компьютере. В частности, формат статей синтаксического словаря, векторная запись словоформ и распределение их по парадигмам, а также формат синтаксических отношений связаны между собой как одно целое. Одной из основных целей диссертационного исследования является построение такого синтаксического анализатора, который позволил бы редактировать синтаксис по ходу исполнения анализатора. Возможность интерактивного редактирования синтаксического словаря и ТСО накладывает дополнительные требования на простоту алгоритмов и программы в целом. Для быстрого обмена данными между различными объектами программы и поиска информации в словаре используются различные коллекции и приемы индексации. В главе четыре показаны различные диалоговые окна и элементы управления, которые присутствуют в интерфейсе программы.

В пятой главе приводится формальное описание *семантики* ЕЯ, показана возможность построения формальной теории семантики и ее компьютерной модели. Модель синтаксиса без модели семантики не представляет практическую ценность как средство общения человека с компьютером. Основными терминами (элементами) семантики являются *понятие*, *признак*, *значение* и *элементарное знание*. Элементарное знание делится на *потенциальное* и *актуальное знания*.

Потенциальное элементарное знание представляет собой последовательность из трех элементов: понятие – признак – значение или $x - w - z$, или $x(w[z])$, где через x , w и z обозначены понятие, признак и значение. Набор *потенциальных элементарных знаний* $x(w[z])$ при фиксированном значении понятия x образуют *понятие* x . Относительно знания $x(w[z])$ можно сказать, что « z есть значение признака w », « z есть значение понятия x », « w есть признак понятия x ». Если тройку элементов $x - w - z$ в понятии x соединить дугами (x, w) и (w, z) , то все тройки, образующие понятие x , будут представлены путями, выходящими из вершины x , а понятие x будет представлено корневым деревом с корнем в x .

Некоторые значения понятия сами могут быть понятиями. Например, **дом – цвет – белый** есть один из путей, образующих понятие с именем **дом**, а в знании **правительство – собственность – дом** именем **дом** обозначено значение понятия с именем **правительство**. В таком случае, понятия, признаки и значения окажутся связанными между собой в единой сети, которая образует *граф семантических отношений* (ГСеО). Термины «понятие x » и «корневое дерево x » в ГСеО используются как синонимы и представляют собой дерево с корнем в x .

Знание составляет предмет семантики. Потенциальное элементарное знание $x(w[z])$ означает *возможность* понятия x обладать значением z . *Актуальное элементарное знание* $x(w[z])$ означает наличие у понятия x значения z . Чтобы отличить два вида знания, актуальное знание обозначается жирным начертанием. Принципиальное отличие потенциального знания $x(w[z])$ от актуального знания **$x(w[z])$** заключается в том, что потенциальные знания **дом(цвет[белый])** и **дом(цвет[красный])** совместимы, а актуальные знания **дом(цвет[белый])** и **дом(цвет[красный])** несовместимы в одном понятии с именем **дом**. Если в понятии $x = \text{дом}$ для признака $w = \text{цвет}$ выбрать одно из возможных значений, например, $z = \text{белый}$, а остальные знания **дом(цвет[v])** отбросить, то этот процесс называется *актуализацией* признака $w = \text{цвет}$ у понятия $x = \text{дом}$.

Предложения языка осуществляют *частичную актуализацию* понятия, то есть актуализацию значений некоторых признаков понятия. Через $x(a, b, \dots, d)$

обозначается понятие x с актуализированными признаками a, b, \dots, d . Через $x(\dots)$ обозначается частично-актуализированное понятие x . Через x обозначается актуализированное понятие x или *вещь* x , в которой актуализированы все признаки понятия x . (Вещь обозначается жирным начертанием) Потенциальное и актуальное знания (понятие и вещь) находятся в таком же отношении как *класс* и *объект* (экземпляр класса) в ООП (объектно-ориентированном программировании). Способы представления понятия, приемы его актуализации и исследования возможности «проникновения» одного понятия в другое путем их частичной актуализации составляют предмет семантики. Изучение этих вопросов может пролить свет на решение принципиальных вопросов компьютерной лингвистики.

Частично актуализированное понятие является понятием. Понятие x называется предком понятия $x(\dots)$. Например, понятие, обозначенное именем **дом**, есть предок понятия, обозначенного именем **белый дом**. Понятие, обозначенное именем **птица**, есть предок понятия, обозначенного именем **гусь**. Отношения типа «предок – потомок» между понятиями называются *родовыми*. Родовые отношения изучаются в *формальной логике*. В диссертации показано, что *вывод силлогизмов* можно свести к поиску путей в заданном графе.

Элементарному знанию **дом – цвет – белый** в синтаксисе русского языка соответствует словосочетания **дом цвета белого, дом белый, дом имеет цвет белый, дом белеет** и т. д. *Отображение* элементов семантики (элементарных знаний) в элементы синтаксиса (словосочетания) можно задавать с помощью *синтаксических форм*, которые были определены в главе 1. В диссертации показано, каким образом синтаксические формы могут быть использованы как операторы, отображающие знание в словосочетание.

Заключение

В диссертационной работе предлагается математическая модель синтаксиса ЕЯ. Отсутствие семантической составляющей у слов и словосочетаний упрощает саму модель и алгоритмы решения задачи СА. Модель синтаксиса определяется

набором синтаксических отношений (ТСО). Семантика в компьютерной модели языка, как правило, учитывается делением лексики на более мелкие части и увеличением числа синтаксических отношений, которые по мере уменьшения их мощности превращаются в семантические отношения. Приближение компьютерной модели языка к ЕЯ путем подобного деления очень усложняет модель языка.

В диссертации предлагается иной подход к построению предложений ЕЯ путем построения двух самостоятельных моделей: синтаксиса и семантики. В первой (второй) модели исследуются и разрабатываются алгоритмы синтаксического (семантического) анализа и синтеза синтаксически (семантически) правильных предложений, а синтаксически и семантически правильное предложение является предложением ЕЯ.

Единственной целью разделения полномочий синтаксиса и семантики является упрощение их моделей. Поставленная цель достигается благодаря тому, что модель синтаксиса (семантики) задается графом синтаксических (семантических) отношений, а синтаксически (семантически) правильным предложением является корневое дерево в рассматриваемом графе.

В диссертационной работе показано насколько простую и наглядную модель имеет синтаксис ЕЯ, а задача СА решается на компьютере за доли секунды.

Работы автора по теме диссертации

1. Шихиев Ф.Ш., Рамалданова Т.О., Формальная теория синтаксиса естественного языка. Сб. статей: Вестник молодых ученых Дагестана №3, Махачкала 2002, с. 69 – 73.
2. Шихиев Ф.Ш., Анализ и синтез словоформ. Сб. статей: Вестник молодых ученых Дагестана №3, Махачкала 2004, с. 108 – 112.
3. Шихиев Ф.Ш., Сетевые алгоритмы построения доказательств в формальной логике. Сб. статей: Вестник молодых ученых Дагестана №3, Махачкала 2004, с. 104 – 108.

4. Шихиев Ф.Ш., Сетевая формулировка задачи синтаксического анализа. /IV Республиканская научно-практическая конференция «Информационные и телекоммуникационные системы: теоретические основы информационных технологий», Махачкала 2005, ДНЦ РАН, с. 124 – 126.
5. Шихиев Ф.Ш., Ясулова Х.С., Модель формального синтаксиса ЕЯ. Сб. статей: Вуз и школа: пути и средства взаимодействия, Махачкала 2005, с. 71 – 72.
6. Шихиев Ф.Ш., Шихиев Ш.Б., Синтаксические операции, категории и парадигмы. Сб. статей: Вуз и школа: пути и средства взаимодействия, Махачкала 2006, с. 67 – 70.