

Московский ордена Ленина, ордена Октябрьской Революции и ордена Трудового Красного Знамени государственный университет имени М.В. Ломоносова

факультет вычислительной математики и кибернетики кафедра алгоритмических языков

Волкова Ирина Анатольевна

АДАПТАЦИЯ И ОБУЧЕНИЕ СИСТЕМЫ ОБЩЕНИЯ С ЭВМ НА
ЕСТЕСТВЕННОМ ЯЗЫКЕ

(01.01.10 — математическое обеспечение
вычислительных машин и систем)

Диссертация
на соискание учёной степени
кандидата физико–математических наук

Научные руководители:

к.ф.–м.н., доцент
Н.П. Трифонов,

к.ф.–м.н., доцент
М.Г. Мальковский

Москва, 1982 г.

Содержание

Введение 2	
Глава 1 . ПРИНЦИПЫ И МЕТОДЫ ОРГАНИЗАЦИИ АДАПТИВНОГО ДИАЛОГА ЧЕЛОВЕКА С ЭВМ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ	8
§ 1. Понятия адаптации и обучения.	8
§ 2. Открытые и базовые знания системы.	9
§ 3. Организация памяти системы	10
§ 4. Метауровень знаний системы о языке.	10
§ 5. Незнакомые и новые языковые объекты.	11
§ 7. Функционирование процедур адаптации и обучения в составе системы TULIPS–2.	12
Глава 2. МОДЕЛЬ РУССКОЙ МОРФОЛОГИИ	14
§ 1. Основные понятия и определения.	15
§ 2. Исключения.	17
§ 3. Чередования.	19
ТАБЛИЦА ЧЕРЕДОВАНИЙ	19
§ 4. Словарь.	23
§ 5. Средства сопровождения словаря системы.	29
§ 6. Процедуры морфологического уровня. Анализ.	31
§ 7. Процедуры морфологического уровня. Синтез.	34
Глава 3. АДАПТАЦИЯ И ОБУЧЕНИЕ СИСТЕМЫ НА ЛЕКСИЧЕСКОМ И МОРФОЛОГИЧЕСКОМ УРОВНЯХ	36
§ 1. Автоматизированное пополнение базы знаний системы о языке.	36
§ 2. Методы и процедуры автоматической адаптации системы к неизвестным ей языковым объектам.	39
Префиксы.	42
Суффиксы	43
Уточнение значений ГП словоформ по последней букве основы (ПБО).	49
Дополнительные правила, используемые при анализе неизвестных слов.	51
§ 3. Обучение системы пользователем.	51
§ 4. Самообучение системы.	52
§ 5. Автоматическое исправление ошибок во входном тексте на естественном языке.	55
Заключение	57
Литература	60
ПРИЛОЖЕНИЕ 1.	62
ПРИЛОЖЕНИЕ 2.	67
ПРИЛОЖЕНИЕ 3.	75

Введение

Привлекательность общения человека с ЭВМ на естественном *языке* не вызывает сомнения. Естественный язык не требует специального изучения, предоставляет широкие возможности для выражения любого содержания, а, следовательно, и является тем средством общения с машиной, которое может удовлетворить пользователей любой специальности и квалификации. В наше время, когда пользователем ЭВМ может оказаться практически каждый (но далеко не каждый захочет изучить какой-либо формальный язык общения с машиной, а в некоторых случаях это будет и нецелесообразно), проблема взаимодействия человека с ЭВМ на естественном языке стала важной практической задачей. Однако при реализации этой проблемы существует много трудностей. Естественный: язык — большая, сложная, постоянно изменяющаяся система, в нем существует множество различных правил, исключений 13 правил (которые в свою очередь являются более частными правилами). Из опыта развития автоматического перевода ясно видно, что поверхностный подход (без выделения семантического, а иногда и синтаксического уровней) к языку не годится. Все ранние системы автоматического машинного перевода оказались неудовлетворительными в основном именно из-за их одноуровневости [1]. Этим единственным уровнем был лексико-морфологический уровень, т.е. некоторое слово входного языка заменялось некоторым переводным эквивалентом — словом выходного языка. Поскольку без обращения к другим языковым уровням обоснованно такой эквивалент словам выбрать невозможно, разобратся в получавшихся переведённых текстах было труднее, чем в непеределённых (фактически это были не переводы, а «подстрочники»). Стало ясно, что удовлетворительного автоматического перевода не может быть без учёта синтаксических и семантических связей слов и фраз, без определённого понимания машиной того, что она переводит. Одни исследователи ограничились таким решением проблемы, когда человек переводит текст, обращаясь за помощью к машине (в этом случае машина выполняет практически роль автоматического словаря); другие — таким, когда машина переводит текст с помощью человека/который вмешивается в её деятельность лишь для преодоления существенных трудностей; а некоторые стали дальше развивать лингвистический аспект машинного перевода (в последние годы, например, во многих системах машинного перевода был значительно развит синтаксический компонент, введено тщательно разработанное синтаксическое представление текста, предприняты попытки введения семантического уровня; но почти все существующие системы машинного перевода в настоящее время либо находятся в начальной стадии разработки, либо используются как экспериментальные системы [2]).

В отличие от работ по машинному переводу, в работах по искусственному интеллекту естественный язык рассматривается менее полно и подробно, но в то же время (а, может быть отчасти, и благодаря этому) практически во всех системах искусственного интеллекта есть уровень семантической обработки» используется некоторое семантическое представление текста, Исследователей в области искусственного интеллекта прежде всего интересует не язык как таковой, не задача составления полного его описания, а задача раскрытия смысла сообщения, написанного на естественном языке, т.е. понимание того, что требует пользователь при очередном обращении к машине. Важно, что в системах искусственного интеллекта (например, в г системе TULIPS-2 М.Г. Мальковского [3], программе, понимающей естественный язык Т. Винограда [4]) раскрытие семантического и прагматического значений сообщения предполагает не только построение описания его смысла, но и выполнение некоторых действий, определяемых этим сообщением — поиск ответа на вопрос пользователя, решение описанной в тексте задачи, преобразование модели «внешнего мира». Смысл, таким образом, трактуется «операционально» — как некоторые действия системы, инициируемые полученным ею сообщением,

Среди экспериментальных систем, имитирующих понимание естественного языка, можно отметить, например, программу М.Г. Мальковского [5], которая решает простейшие арифметические задачи, сформулированные на русском языке. Описание языка в этой программе включает описание слов и конструкций, значимых в рассматриваемой проблемной области, т.е. соответствующих «объектам счёта», их «владельцам», отношениям, действиям и обстоятельствам. При анализе очередной фразы программа выделяет такие конструкции и заносит полученную информацию в таблицу, отображающую условие задачи. Нужные действия для решения задачи программа определяет по семантическим характеристикам языковых объектов, например, по семантическому типу глаголов («плюс» или «минус»). При обработке неполных предложений используется информация, полученная при анализе соседних фраз условия.

Семантическая система программы, понимающей естественный язык, Т. Винограда построена на базе группы программ, которые являются экспертами по отношению к конкретным синтаксическим структурам. Они рассматривают как структуры, так и значения слов, и строят выражения на внутреннем языке, которые будут использоваться при ответах на вопросы и выполнении дедуктивных выводов [4]. В системе Т. Винограда предполагается, что разговор идёт с роботом, имеющим «руку» и «глаз» и способным манипулировать игрушечными кубиками на столе. Робот отвечает на вопросы и воспринимает информацию, которая ему в дальнейшем может понадобиться.

Другой пример программы, понимающей естественный язык — программа MARCIE, созданная под руководством Р. Шенка. Эта программа производит умозаключения и осуществляет перефразирование предложений на английском языке. В основе MARCIE — представление смысла фраз естественного языка в терминах концептуальной зависимости [6]. Под теорией концептуальной зависимости Р. Шенк понимает теорию того, как человек обрабатывает фразы на естественном языке. Проблема обработки фраз рассматривается в трёх аспектах: отображение предложений в их смысловое представление; хранение в памяти и осуществление умозаключений относительно полученного смысла, и перевод смыслового представления на естественный язык.

В последнее время появилось много экспериментальных и прикладных систем, обеспечивающих диалог человека с ЭВМ на так называемом «ограниченном» естественном языке. Наиболее богатыми по своим возможностям среди таких систем являются системы ПОЭТ Э.В. Попова [7], ДИЛОС В.М. Брябрина и других [8], ДИСПУТ Л.И. Микулича и А.Я. Лервоненкиса [9].

Система ПОЭТ программа обработки экономических текстов: вопросно–ответная система, общение с которой ведётся на ограниченном русском языке. Система рассчитана на работу со словарём в несколько тысяч слов и большой базой данных. Она состоит из следующих компонентов: языка общения, словаря, семантической сети, базы данных и комплекса программ, которые выполняют морфологический, синтаксический, семантический анализ семантическую интерпретацию и формирование ответа, синтаксический и морфологический синтез С. Знания в системе распределены между словарём, семантической сетью и базой данных. В словаре хранится информация об основах слов и словосочетаний.

Система ДИЛОС — с одной стороны экспериментальный комплекс для отработки различных идей и методов представления знаний и обработки естественного языка, с другой — система, реализованная в виде законченного набора программ и используемая рядом коллективов с целью решения практических задач. Все процедуры ДИЛОС сгруппированы в несколько блоков, называешь «процессорами»: лингвистический, семантический, логический, информационно–поисковый и вычислительный. Работа этих процессо-

ров обеспечивает накопление, поиск информации в базе данных системы, трансляцию и счёт прикладных программ, логический анализ моделей проблемных областей с целью соотнесения с ними сообщений и вопросов пользователя.

В основе системы ДИЛОС лежит специально разработанный формальный язык (Ф–язык), который является средством представления знаний о различных проблемных областях, внутреннего отображения смысла входных естественно–языковых запросов, инструментом для обработки Ф–объектов. Лингвистический процессор системы осуществляет преобразование входных фраз с естественного языка на Ф–язык. При каждом сеансе работы с пользователем происходит: считывание текста, поступившего от пользователя, диалог с ним с целью определения «незнакомых» системе слов, генерация Ф–выражений, являющихся интерпретацией входного естественно–языкового текста, обращение к пользователю с целью выяснения правильности построенных Ф–выражений (это предполагает знание пользователем Ф–языка!), семантическая интерпретация и исполнение построенных ф–выражений. Отметим, что по желанию пользователя вновь определённые слова запоминаются системой, и таким образом, происходит элементарное обучение системы пользователем, но для этого, как отмечалось, от пользователя требуются специальные знания.

Специализированная диалоговая система ДИСПУТ входит в состав автоматизированной системы слежения за контейнерными перевозками и предназначена для получения информации из базы данных и решения некоторых задач оперативного управления. Система ориентирована на пользователя, не являющегося программистом. Запросы к системе формулируются на естественном (русском) языке с лексикой, ограниченной тематикой задачи, и фиксированными семантикой и прагматикой. В системе ДИСПУТ нет средств для синтеза развёрнутого ответа на естественном языке, практически совсем не рассматривается морфологический уровень. Используются относительно простые входные и выходные сообщения фиксированного формата, ответы системы могут также генерироваться в виде таблиц. Приемлемость такой организации диалога с пользователем определяется узостью проблемной области, в которой функционирует система.

Практически единственным достоинством систем, обеспечивающих диалог с ЭВМ на ограниченном естественном языке, является хорошая «читабельность» текстов, написанных на их входном языке. Но ограниченный естественный язык не предоставляет всех удобств, которые пользователь имел бы при общении с ЭВМ на действительно естественном языке — языке, употребляемом человеком в его повседневной речевой практике: ограниченный естественный язык требует предварительного изучения; при диалоге пользователь должен строго придерживаться принятых в системе ограничений на грамматику и словарь, что существенно затрудняет общение с машиной. Во многих ситуациях выучить некоторый несложный формальный язык и общаться с машиной на нём пользователю даже проще, чем постоянно следить за соблюдением ограничений на используемый повседневно и произвольно естественный язык. Ответим, что в этих системах как правило нет средств, которые позволили бы гол самостоятельно разобраться в незнакомой ситуации: определить нужные характеристики роль а тексте неизвестных системе слов, понять «нестандартные» структуры предложений, обнаружить и исправить ошибки в тексте запомнить новые факты о языке для использования их в дальнейшем (за исключением простейших средств обучения пользователя например, в ДИЛОС'е).

Желательно, чтобы в системах, общающихся с человеком на естественном языке, на входной текст накладывались ограничения более «мягкие» и незаметные для пользователя, чтобы в их составе были средства адаптации к незнакомым языковым объектам и обучения языку, обеспечивающие «разумную» реакцию на незнакомый языковой объект (боль-

шинство существующих систем при появлении незнакомого объекта либо отказывается выполнять какие-либо действия, либо просит дать синоним незнакомого слова, что бывает затруднительно, либо просто игнорируем незнакомый языковой объект). В большей степени это требование касается лексики как наиболее подвижной части естественного языка. В речевой практике человека постоянно встречаются различные вариантные, просторечные и даже заведомо неграмматичные формы слов, возможность появления которых также следует учитывать.

Настоящая работа посвящена разработке программных сред позволяющих сделать диалог с ЭВМ на естественном языке естественным и необременительным. Это средства адаптации к незнакомым языковым объектам и средства обучения (запоминание нового для использования в дальнейшем), которые позволяют пользователю употреблять любые незнакомые системе слова, словосочетания, синтаксические конструкции, вводить свои обозначения и сокращения. Адаптация и обучение рассматриваются только на лексическом и морфологическом уровнях. Это объясняется как сложностью и неизученностью проблемы, так и отсутствием полных, достаточно формализованных лингвистических знаний о языке на других языковых уровнях (синтаксическом, семантическом).

По-видимому, первой экспериментальной системой, в которой результаты исследования проблем адаптации и обучения при диалоге с ЭВМ на естественном языке получили достаточно полное воплощение, является система TULIPS М.Г. Мальковского [10], которая представляет собой обучаемый решатель задач (арифметических и «бытовых»), сформулированных на русском языке. Адаптация и обучение в этой системе базируются на обширных начальных знаниях о языке. Предполагается, что большинство нужных слов, синтаксических конструкций система уже знает, и работа средств адаптации и обучения, а также возможные при этом обращения за помощью к пользователю, происходят крайне редко (чтобы не обременять пользователя).

Методы обучения языку исследуются также при работе над экспериментальной системой, разрабатываемой под руководством В.Н. Гладуна [11] но принципы обучения там совсем другие. Обучение в этой системе начинается с «нуля», т.е. в начале своей работы всё, что система встречает, ей незнакомо. Способность выполнять некоторые преобразования текста (понимать текст) система приобретает в результате обучения языку на примерах. Процесс обучения обеспечивает адаптацию системы к профессиональным языкам и различным особенностям языка. Выделенная на этом этапе обучения информация используется при анализе и синтезе текстов вместо категорий и правил грамматики естественного языка. Этот принцип обучения языку во многом перекликается с принципом построения лингвистических дешифровочных моделей, в которых рассматриваются процедуры, предназначенные для обнаружения того или иного языкового явления, а исходной информацией является только текст (т.е. все сведения о языке извлекаются только из текста). Такой подход может оказаться непригодным при практическом применении обучаемых систем из-за затраты большого количества машинного времени на адаптацию и обучение и, возможно, слишком частых обращений за разъяснениями к пользователю.

Дальнейшее развитие методы адаптации и обучения получили в новой версии системы TULIPS — системе TULIPS-2 [3], для которой и реализованы предлагаемые в данной работе алгоритмы и процедуры. TULIPS-2 — экспериментальная система, но исследуются в ней методы и алгоритмы могут быть применимы (и предполагается, что эти средства именно в такой форме будут нужны) и в будущих практических системах.

Средства адаптации и обучения на лексическом и морфологическом уровнях, описанные в работе, разработаны с ориентацией на морфологическую модель русского языка системы TULIPS-2 (см. гл. 2), созданной на базе фундаментальных лингвистических исследований

[12]. Известны также морфологический анализатор Г.Г. Белоногова [13], модель русской морфологии И.А. Большакова [14].

Морфологический анализ слов в системе Г.Г. Белоногова происходит на основе словаря словоформ. Задача получения грамматических и семантических признаков слов сводится в основном к поиску в словаре, и только в тех случаях, когда данной словоформы там нет, происходит морфологический анализ слова. Применение словаря словоформ позволяет преодолеть трудности, связанные с такими явлениями словоизменения и словообразования, как чередования, наличие супплетивных форм слов. Это достигается путём отображения в словаре парадигматических связей между словоформами независимо от их буквенного оформления. Но отсутствие средств распознавания указанных явлений сказывается на «новых» (не записанных в словаре) словах. В системе также нет средств уточнения характеристик новых слов в случае неверного их определения. Помимо словаря словоформ используются также словарь пятибуквенных концов слов (для определения грамматических признаков новых слов), словари окончаний и сочетаний суффиксов (для установления парадигматических связей на уровне основ слов и на уровне «семантических эквивалентов слов»: «завод» — «заводской»). Всем элементам словарей ставится в соответствие некоторый числовой код. Новому слову присваиваются грамматические признаки того элемента словаря пятибуквенных концов слов, который более близок к его последние пяти буквам по численному значению. Аналогично анализатор работает и при определении семантических признаков слова. При этом точность анализа новых слов достаточно высокая (при словаре словоформ объёмом в сто и более единиц — около 90%, при пустом словаре словоформ — не более 78% [13]). Реализованная система содержит словари: словоформ на 24 000 единиц, пятибуквенных концов слов на 8715 единиц, окончаний на 106 единиц, сочетаний суффиксов на 1078 единиц.

Морфологическая модель И.А. Большакова отличается полнотой и подробностью описания правил морфологического синтеза русских словоформ (в настоящее время полностью разработан только морфологический синтезатор русского языка). В ней учтены все тонкости русского словоизменения, вплоть до правил, справедливых лишь для нескольких слов. Это — строгая, хорошая лингвистическая модель. Недостатком модели является то, что она неудобна для практического применения в силу громоздкости описаний морфологических признаков слов и того, что процесс синтеза словоформ занимает слишком много машинного времени; даже в простейших случаях для генерации словоизменительной парадигмы, например, существительного требуется работа всех шести групп правил перехода от одного уровня представления словоформы к другому (рассматриваются семь морфологических уровней представления словоформы) и нескольких десятков конкретных правил. Модель русской морфологии представлена в виде формальной грамматики, и фактически каждая словоформа парадигмы порождается из некоторого начального символа в результате применения определённого множества правил подстановки.

Одним из важных аспектов адаптации и обучения, рассматриваемых в данной работе, который не принимался во внимание ни в одной из существующих систем общения с ЭВМ на естественном языке, является учёт индивидуальных речевых особенностей пользователя. Основное его назначение — позволить пользователю в ходе общения с машиной определить «свой язык», ввести любые сокращения и обозначения. Учёт индивидуальных речевых особенностей предусматривает также запоминание в специальных участках памяти системы свойственных пользователю отклонений от общелитературных норм («лектора»), отличных от используемых системой правил морфологии и синтаксиса (например, употребление прилагательного после определяемого существительного: «Пришли в лес большой», в нарушение стандартного и известного системе порядка слов — «Пришли в большой лес»). Всё это значительно облегчает диалог человека с машиной: во-первых, при общении с системой пользователь может использовать язык, употребляемый им в обыч-

ной речевой практике, во-вторых, повторное появление свойственных для пользователя неграмматичных вариантов слов уже не требует работы всего аппарата распознавания таких форм слов, а значения их грамматических характеристик считывается непосредственно из зоны пользователя (см. гл. 1, § 3). Также важно автоматическое исправление ошибок во входном тексте, что экономит время как пользователя, так и машины.

В главе 1 описаны основные методы и принципы организации адаптивного диалога (в указанном выше смысле) человека с ЭВМ на естественном языке» а также методы представления знаний о языке, позволяющие обеспечить адаптацию и обучение системы. В главе 2 приводится модель русской морфологии, используемая в системе TULIRS-2, описывается словарь системы и процедуры морфологического уровня, В главе 3 — на лексическом и морфологическом уровнях — рассмотрены методы, алгоритмы и процедуры автоматизированного пополнения базы знаний системы о языке, адаптации к незнакомым ей языковым объектам и индивидуальным речевым особенностям пользователя, обучения системы естественному языку и автоматического исправления ошибок во входном тексте. В приложениях приводятся таблицы, задающие деление слов каждого морфологического класса на парадигматические классы, таблицы соответствия словарной информации из словаря А.А. Зализняка словарной информации системы TULIPS-2, а также правила работы с процедурой автоматизированного пополнения словаря системы КОД.

Глава 1 . ПРИНЦИПЫ И МЕТОДЫ ОРГАНИЗАЦИИ АДАПТИВНОГО ДИАЛОГА ЧЕЛОВЕКА С ЭВМ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

Адаптивный диалог на естественном языке* это диалог, устойчивый к различного рода отклонениям пользователя от знаний системы о языке (ЯЗС), т.е. от имеющихся в ней описаний языка общения, случайным ошибкам. Реализуется он с помощью средств адаптации и обучения, соответствующей организации базы знаний системы, метауровня ЯЗС, широкого использования предсказаний о различных характеристиках анализируемого языкового объекта (синтаксический класс, грамматические признаки, семантическое значение).

Предлагаемые в этой главе принципы и методы излагаются применительно к системе TULIPS-2, хотя могут быть использованы в других системах, обеспечивающих диалог человека с машиной на естественном языке.

§ 1. *Понятия адаптации и обучения.*

Автоматическая адаптация (на языковом уровне) — это способность системы установить характеристики и роль в тексте впервые встретившихся ей языковых объектов — форм не представленных в словаре слов, незнакомых словосочетаний, не описываемых грамматикой синтаксических конструкций; т.е. это действительно адаптация, приспособление ЯЗС (описаний языка) к незнакомым (не входящим в это описание) языковым объектам.

Обучение — это запоминание результатов автоматической адаптации (изменений и дополнений ЯЗС) или явных указаний пользователя (например, его ответов на вопросы системы) для использования их в дальнейшем. В зависимости от того, кто является инициатором запоминания нового, различаются самообучение и обучение системы пользователем. При обучении пользователем система (формируя при этом соответствующие описания языковых объектов) запоминает только то, что пользователь явно просит запомнить, например, вводимые им обозначения, новые термины. При самообучении формирование описаний новых языковых объектов и запоминание их иницируется не непосредственными указаниями пользователя, а общими языковыми знаниями системы.

Механизмы автоматической адаптации работают параллельно с анализом текста, обучение же, как правило, происходит в конце сеанса работы системы, т.е. работы очередного пользователя в очередной проблемной области. Это объясняется различием целей адаптации и обучения. Если основная функция адаптации — обеспечить завершение анализа очередного сообщения, то результатом обучения должно быть формирование и запоминание достаточно полного и точного описания ранее незнакомого объекта. Уточнение описания объекта может происходить во время сеанса, при появлении других его форм, удовлетворяющих описанию и в то же время позволяющих уточнить некоторые его детали. По окончании же сеанса производится тщательный (и, следовательно, требующий значительного времени) анализ подготовленных к запоминанию описаний. Определяется область запоминания — зона пользователя, зона проблемной области или общая зона (см. § 3); уточняются грамматические характеристики объекта; объединяются словарные статьи с тождественными входами. Эти два фактора — возможность уточнения характерис-

* Термин «адаптивный диалог» употребляется и в несколько других смыслах [5][16], и не обязательно в связи с использованием для диалога с машиной естественного языка.

тик объекта в ходе сеанса и длительность последующей подготовка к запоминанию — и объясняют «отсроченность» этапа переноса новых ЯЗС в долговременную память системы. В течение сеанса сформированные предварительные описания хранятся в активной памяти системы в особой форме (см. § 5).

На лексическом и морфологическом уровнях наличие средств, реализующих адаптацию и обучение системы, позволяет выявлять и учитывать;

- 1) новые общезначимые языковые факты (новые слова);
- 2) специфику языка проблемной области (специфические термины, профессионализмы);
- 3) специфику языка пользователя (вплоть до отклонений от общелитературных норм — «шофера»);
- 4) вводимые пользователем обозначения и сокращения;
- 5) случайные ошибки (перестановка, замена букв и др.).

§ 2. Открытые и базовые знания системы.

На всех языковых уровнях ЯЗС делятся на «открытые» и базовые. Открытыми называется знания, которые могут пополняться в ходе диалога по инициативе пользователя или системы. Базовая часть ЯЗС формируется заранее лицами, обслуживающими систему, по материалам обработки характерных текстов или опроса пользователей. Открытым является множество описаний слов и словосочетаний русского языка, множество описаний различных отклонений конкретных пользователей от используемой системой модели языка. Базовые ЯЗС на лексическом и морфологическом уровнях включают: полное описание русского словоизменения, описание отдельных способов словообразования, различных типов речевых ошибок пользователей, базовый словарь основ и неизменяемых слов, словари служебных морфем (флексии, отдельные суффиксы и префиксы), т.е. базовые ЯЗС включают необходимые минимальные знания системы и знания о языковых объектах и конструкциях, появление в тексте которых можно предусмотреть заранее. Достаточно обширными должны быть ЯЗС и на других языковых уровнях.

Таким образом, система начинает учиться языку не с «нулевого», а с достаточно высокого начального уровня знаний. А адаптация и обучение рассматриваются прежде всего с позиций практического удобства работы с системой, необременительности диалога для пользователя. Они направлены на обеспечение и повышение самостоятельности работы системы в незнакомых ситуациях и повышение эффективности её работы в знакомых (например, за счёт преобразования описаний часто встречающихся языковых объектов в форму, обеспечивающую более быстрый анализ их при повторном появлении в тексте). Этот метод — формирование «шаблонов» — активно используется при обучении на всех языковых уровнях (в первую очередь на синтаксическом). На морфологическом уровне такими шаблонами являются, например, описания неизменяемых слов (см. гл. 2, § 4), которые позволяют анализировать входные словоформы во много (в 10–20) раз быстрее, чем при использовании описаний основ. Шаблоны этого типа генерируются системой для конкретных, часто встречающихся во входном тексте форм известных изменяемых слов.

§ 3. Организация памяти системы*

Память системы TULIPS–2 организована с учётом того, что система макет используется в различных проблемных областях, в каждой из которых с ней будут общаться несколько пользователей.

Память системы делится на три уровня: долговременная память, содержащая всю совокупность знаний системы; активная память, формируемая в начале очередного сеанса работы; и оперативная память, в которой производятся все преобразования содержимого памяти во время работы системы над очередной задачей (т.е. там хранятся некоторые промежуточные результаты, в которых значения каких-либо характеристик могут уточняться в дальнейшем; например, фрагменты словарных статей новых слов, полученные в результате автоматической адаптации). В долговременной памяти системы выделено несколько зон: общая, где представлены знания, которые могут использоваться более, чем одним пользователем, и (или) более, чем в одной проблемной области; частные, где описываются языковые особенности пользователей или специфика употребления языка в каждой из известных системе проблемных областей, например, характерные термины. В начале КАЖДОГО сеанса пользователь называет себя и указывает область работы. По этим данным начинается формирование активной памяти системы — активация. Процесс активации реализует настройку системы на очередной сеанс. При этом из общей зоны и зон текущих проблемной области и пользователя считываются и переводятся в нужную форму только те знания, которые в данном сеансе могут оказаться уместными. Уместность ЯЗС определяется по специальным «указателям уместности» — числам, характеризующим целесообразность использования того или иного описания языкового объекта в данном сеансе, порядок перебора вариантов словарных значений и правил грамматики или вызова процедур. В зависимости от частоты встречаемости конкретных языковых объектов в текстах, обрабатываемых системой во время сеанса, их указатели уместности могут автоматически изменяться. Достижение некоторого «порогового» значения указателя уместности описания может служить сигналом для формирования упоминавшегося выше «шаблона». Изменения указателей уместности выполняют общие механизмы преобразования памяти системы, не рассматриваемые в данной работе. Настройка на сеанс позволяет не только сократить объем описаний, просматриваемых при анализе текста, но и учесть специфику языка пользователя и проблемной области, поскольку специфические описания — синтаксические правила, семантические значения, словари — при анализе просматриваются в первую очередь.

§ 4. Метауровень знаний системы о языке.

Название метауровень ЯЗС принято потому, что на данном уровне знания системы включают в себя не описания самого естественного языка, а описания некоторых возможных отклонений от ЯЗС системы, типов реакции на них в каждом конкретном случае, т.е. описание ЯЗС и их преобразований.

Метауровень ЯЗС представляется в форме процедур, описывающих реакцию системы на появление в тексте незнакомых ей языковых объектов [18]. Эти процедуры совместно с процедурами адаптации устанавливают тип объекта (искажённый вариант известной языковой единицы или впервые встретившаяся правильная языковая форма), его роль в тексте, подготавливают информацию об объекте или факте его появления для запоминания. Учитывая результаты обращения к метауровню, система либо корректирует и пополняет

* См. также [17][18].

свои общие знания (в случае новых слов), либо запоминает специфичные конструкции (просторечные варианты, профессионализмы) в зоне пользователя, либо просто «пропускает» неграмматичный вариант на последующие этапы анализа, если характер допущенной ошибки и функции неверного варианта ясны, например, ошибку удаётся исправить с помощью блока нейтрализации случайных ошибок (см. гл. 3, § 4 стр. 52). Система может послать пользователю сообщение об ошибке или обратиться к нему за разъяснением.

Отметим, что определить в каждом конкретном случае тип незнакомого языкового объекта очень сложно (особенно из-за того, что система ориентирована на работу с открытым словарём). В данной работе не ставилась цель подробно описать метауровень для русского языка (даже на лексическом и морфологическом уровнях, которые находятся в центре внимания); рассматриваются лишь основные принципы и отдельные методы его использования. Принципиально важно, что знания метауровня позволяют избегать слишком частых обращений к пользователю за разъяснениями. В то же время возможность такого обращения гарантирует работоспособность систем и в тех ситуациях, когда знаний метауровня недостаточно для автоматического определения типа незнакомого объекта (метауровень позволяет найти конфликт, а разрешение его предусматривает помощь пользователя).

§ 5. Незнакомые и новые языковые объекты.

В системе различаются знакомые и новые языковые объекты. Языковой объект считается знакомым, если он встретился системе впервые, т.е. не может быть воспринят ею как объект (или форма объекта), описание которого считано в долговременную память системы. С помощью средств адаптации формируется предварительное описание такого объекта, например, словарная статья для впервые встретившегося слова. Сам объект рассматривается как «новый», а соответствующее описание получает специальный признак «описание нового объекта». Наличие этого признака позволяет системе самостоятельно уточнять построенные описания при появлении в тексте других форм или конструкций, удовлетворяющих этим описаниям, причём без тех дополнительных предположений, которые предусмотрены метауровнем в подобных случаях для знакомых слов (см. гл. 3, § 4 стр. 52).

На лексическом уровне каждое знакомое слово после его морфологического анализа получает признак «новое слово» и его описание заносится в словарь новых слов. Обычно этот признак снимается в конце сеанса работы системы, когда все нужные характеристики слова выяснены, и оно записывается в долговременную память системы. При этом также работают процедуры уточнения грамматических характеристик новых слов, обобщения словарных статей новых слов с одинаковой основой («жар», «жарить»), автоматического исправления возможных ошибок в основах (см. гл. 3, § 3-4 стр. 51, 52). Например, появление формы винительного падежа существительного мужского рода позволяет уточнить значение категории одушевленности. Так, при первом появлении существительных «киоск» и «мальчик» во фразе «у метро стоит киоск / стоит мальчик» об их одушевлённости ничего сказать нельзя. Но при повторном появлении этих существительных, например, в таком контексте: «увидев этот киоск / этого мальчика». Система однозначно определяет, что «киоск» — неодушевлённое, существительное, а «мальчик» — одушевлённое.

§ 6. Предсказания*

Предсказания — это формируемые в ходе синтаксического и семантического анализа отдельного предложения или всего входного текста ожидаемые характеристики незнакомого языкового объекта.

При автоматической адаптации источниками информации о незнакомом языковом объекте являются: структура объекта, контекст, общие языковые знания системы. Использование предсказаний при анализе очередного сообщения — основной способ учёта контекста. Для слов могут предсказываться синтаксический класс, некоторые значения грамматических характеристик (число, падеж, наклонение — аналог времени глагола — и другие), основа, семантическое значение слова. Аппарат предсказаний обеспечивает необходимую гибкость управления анализом текста и, главное, возможность автоматической обработки незнакомых языковых объектов. Поскольку в системе используются обширные базовые знания о языке, то формируемые предсказания в большинстве случаев являются достаточно определёнными и точными. С помощью предсказаний можно правильно понять незнакомые слова с особенностями изменения, например, воспринять словоформу «глаз» как существительное в форме множественного числа родительного падежа, хотя по общим правилам в этой форме слова должна была бы быть флексия «-ов» («алмазов»); обнаружить ошибку, например, в слове «лекторá», хотя это слово и имеет флексию «-а» в форме единственного числа родительного падежа. Для слов, анализируемых без предсказаний, могут выдаваться либо все варианты анализа, либо только наиболее вероятные (см. § 7, стр. 12). Если же незнакомое слово анализируется с предсказаниями, то результатом являются либо только предсказанные, либо все допустимые варианты анализа. Как правило, не включённые в результат варианты сохраняются и могут быть получены при повторной обработке слова, например, с новыми предсказаниями. Режим работы устанавливается процедурами, охватывающими морфологический анализатор.

§ 7. Функционирование процедур адаптации и обучения в составе системы TULIPS-2.

Анализ входного сообщения системы непосредственно включается в целостный процесс взаимодействия системы с пользователем, т.е. смысл сообщения (конечная цель анализа) определяется с учётом текущей области работы, характерных для неё типов заданий и объектов, известных системе языковых особенностей пользователя. На любом этапе анализа управление может быть передано любому блоку системы, способному обработать выявленную информацию для устранить возникший конфликт (в частности блоку адаптации и обучения системы). В отдельных случаях возможно обращение за разъяснениями к пользователю. Таким образом, работа аппарата адаптации и обучения включается в процесс анализа входного текста. Соответствующие процедуры вызываются при возникновении «конфликтных» ситуаций. Работа процедур адаптации и обучения проходит в рабочем режиме, т.е. непосредственно в ходе диалога с пользователем. При этом в ЯЗС системы (временно для с последующим переносом в долговременную память) добавляются описания незнакомых ранее языковых объектов, в том числе: объектов, специфичных для речи отдельных пользователей, вводимых в диалоге обозначений. Кроме рабочего режима при работе системы используется также режим «пополнения», когда лицо, обслуживающее систему (или же пользователь) с помощью процедур автоматизированного пополнения словарей и грамматик (см. гл. 3, § 1, стр. 36) предварительно формирует базовые ЯЗС, например, при использовании системы в новой проблемной области.

* См. также [19].

Знания системы о русском языке связаны взаимными ссылками со знаниям о внешнем мире, пользователе. Анализ текста ведётся с учётом этих связей, причём определяющим считается уровень содержания — «опора на высшие аспекты значения». Отметим, что знания системы о языке могут представляться в системе в виде описаний и в виде процедур. В виде плэнерских списков свойств [20], например, в системе хранятся описания флексий русского языка, основ и неизменяемых слов (см. гл. 2, § 4, стр. 23). Чередования русского языка, знания метауровня и некоторые другие представлены в системе процедурно, т.е. в виде программ, выполняющих действия, соответствующие этим знаниям (см. гл. 2, § 6, стр. 31). Безусловно, при использовании и реализации удобнее, чтобы процедуры, перерабатывающие знания о языке, зависели только от формы их описания, а не от содержания соответствующих данных ([21], стр. 20). Процедурное представление знаний о языке в системе TULIPS-2 не означает отступления от этого принципа, и вызвано спецификой языка реализации — языка ПЛЭНЕР-БЭСМ. Предоставляя широкие возможности для обработки различных списочных структур, работы со списками свойств и базами данных, он в то же время «навязывает» и характерный для него процедурный способ представления знаний (так как процедура является основной единицей языка, в виде процедур оформляются любые действия программы).

Важно также, что система TULIPS-2 является многовариантной системой. Так, например, при анализе незнакомых системе слов в качестве результата анализа сначала выбирается наиболее вероятный его вариант (с учётом конкретного пользователя, проблемной области и предсказаний). Но всегда предусматривается возможность повторной обработки слова (например, с другими предсказаниями). Непредсказанные варианты анализа запоминаются системой (см. гл. 2, § 6, стр. 31).

Например, в проблемной области «Арифметические задачи» при анализе словосочетания «один мальчик» более вероятна интерпретация «мальчик → количество == 1», чем «мальчик → один/первый или некоторый», хотя и возможно условие задачи, соответствующее второй интерпретации: «Один мальчик купил 3 книги, другой — 7. Сколько книг купили мальчики?». Если анализ условия этой задачи начинается с первой фразы (одна из возможных стратегий анализа), то, хотя первоначально будет выбран первый из указанных вариантов интерпретации слова «один», затем будет получен и второй (в данном случае правильный).

Глава 2. МОДЕЛЬ РУССКОЙ МОРФОЛОГИИ

Предлагаемая модель русской морфологии, используемая в системе TULIPS–2, разработана на основе морфологической модели системы TULIPS–2 [10] и данных, содержащихся в «Грамматическом словаре русского языка» А.А. Зализняка [12].

Хотя ряд методов построения морфологической модели системы TULIPS (использование словаря основ и неизменяемых слов, непосредственная «погруженность» морфологического анализа в синтаксический и семантический анализ текста, учёт предсказаний) и «перенесён» на морфологическую модель системы TULIPS–2, последняя является значительно более полной и адекватно. В ней уточнено разбиение слов на классы, учитываются различные типы исключений. Очень существенно, что при анализе незнакомых слов стала учитываться структура основы (что заметно повысило точность их анализа). Более полной является также программная реализация модели (реализованы морфологический синтез словоформ, средства сопровождения словаря системы).

Существенно отличается предлагаемая морфологическая модель и от модели, представленной в словаре А.А. Зализняка. Так, словарь А.А. Зализняка представляет собой словарь начальных форм слов, а морфологическая модель предусматривает только синтез словоформ, записанных в словаре. Морфологическая модель системы TULIPS–2 — программно реализованная модель. В ней, как отмечалось выше, используется словарь основ и неизменяемых слов, разработаны средства как синтеза, так и анализа словоформ, средства обработки и восстановления словарной информации незнакомых слов. В модели системы TULIPS–2 изменено предлагаемое в словаре А.А. Зализняка разбиение слов на парадигматические классы, уточнены некоторые особенности словоизменения. Так, например, в словаре А.А. Зализняка наличие любого чередования в словах субстантивного, адъективного и местоимённого склонения обозначается символом «*», а у глаголов наличие чередований (51 тип!) явно совсем не указывается.

В модели системы TULIPS–2 изменён состав глагольной парадигмы: в неё не входят формы причастий, поскольку они являются изменяемыми формами (склоняются). Парадигма причастий совпадает с парадигмой слов адъективного склонения (имеющих соответствующий тип склонения). «Дополнительные» (по сравнению со словами адъективного склонения) характеристики причастий (возвратность, вид, время, залог) также указываются в словарных статьях (см. § 4, стр. 23).

В то же время ориентация на словарь А.А. Зализняка в системе TULIPS–2 с одной стороны гарантирует полноту описания русского словоизменения, с другой — существенно облегчает процесс формирования словаря системы, поскольку используемые в модели морфологические признаки слов однозначно определяются по признакам, приводимым в этом словаре (см. приложение 2, стр. 67).

Описываемая модель русской морфологии программно реализована на языке ПЛЭНЕР–БЭСМ [20] для диалоговой системы искусственного интеллекта TULIPS–2. В принципе модель может работать и независимо от ней в других системах автоматической обработки текстов на естественном языке. Например, модель уже используется в лаборатории по применению вычислительных средств в гуманитарных науках при решении некоторых задач прикладной лингвистики (для этого ряд программ был переписан на языке ПЛ/1 для ЕС–1022).

§ 1. Основные понятия и определения.

В модели множество слов русского языка разбивается на два основных класса — неизменяемые (Н-слова) и изменяемые, т.е. склоняемые или спрягаемые (И-слова). Совокупность форм И-слова (словоформ) образует его парадигму. В каждой словоформе можно выделить основу и окончание, или флексию (возможно пустую — «-∅»), соответствующую конкретной форме И-слова; за флексией может следовать постфикс, например, возвратная частица «-ся»/»-сь». Учитываются также слова с дефисом. Если у такого слова изменяется только одна часть («тёмно-синий»), то проблем никаких не возникает (дефис рассматривается как часть основы); если же изменяются обе части («Баба-Яга»), то у слова выделяются две основы и две флексии (а, возможно, и постфиксы).

С основой И-слова, Н-словом, флексией и словоформой связывается описание значения соответствующего объекта, включающее описание его грамматических характеристик; лексических связей — синонимы, производные слова; семантического и прагматического значения. Грамматические характеристики (для флексий указываются только они) определяют сочетаемость основ и флексий и синтаксические признаки объектов всех четырёх типов. К грамматическим характеристикам морфологического уровня относятся морфологический (словоизменяемый) класс — М-класс, парадигматический класс — П-класс, чередования и исключения. Синтаксическим показателем является синтаксический класс — С-класс. Как словоизменяемые, так и синтаксические признаки определяются набором значений грамматических переменных. Грамматическая переменная (ГП) — переменная одного из следующих типов: одушевлённость, род, число, падеж, вид, лицо, залог, возвратность, наклонение, время, степень — принимает закодированное целым числом значение из некоторого множества допустимых, а именно:

ГП	Значение ГП	Код значения ГП
одушевлённость	одушевлённость	1
	неодушевлённость	2
род	мужской	1
	женский	2
	средний	3
число	единственное	1
	множественное	2
падеж	именительный	1
	родительный	2
	дательный	3
	винительный	4
	творительный	5
	предложный	6
	второй родительный	7
	второй предложный	8
вид	совершенный	1
	несовершенный	2
лицо	первое	1
	второе	2
	третье	3
залог	действительный	1
	страдательный	2
возвратность	возвратность	1
	невозвратность	2
время	настоящее	1

ГП	Значение ГП	Код значения ГП
	прошедшее	3
наклонение или наклонение–время	настоящее	1
	будущее	2
	прошедшее	3
	сослагательное	4
	повелительное	5
	инфинитив	6
степень (сравнительная)	сильная	1
	слабая	2

Если значение ГП неопределённо, указывается список возможных значений; «0», по соглашению, обозначает любое допустимое значение ГП.

Совокупность ГП, по которым изменяется И–слово («свободных ГП»), определяет его парадигму, а спектр значений этих переменных — число элементов парадигмы.

М–класс — множество И–слов, имеющих: 1) общий набор ГП; 2) общий набор свободных ГП; 3) общий спектр значений ГП. Хотя деление на М–классы не совпадает с делением на части речи, 7–й класс образован в основном существительными, 8–й — прилагательными, 9–й — глаголами. В модели рассматриваются три класса склоняемых И–слов: местоименные (класс №5), субстантивные (класс №7), адъективные (класс №8) и один класс спрягаемых (класс №9), ГП, свободные ГП, спектр значений свободных ГП для каждого М–класса приводятся ниже.

М–класс	С–класс	ГП	Свободные ГП	Спектр значений свободных ГП	Примеры
5	5	одушевлённость, род, число, падеж	род, число, падеж	6 падежей ед. числа мужского рода 6 падежей ед. числа женского рода 6 падежей ед. числа среднего рода 6 падежей мн. числа	
7	7	одушевлённость, род, число, падеж	число, падеж	6 падежей (традиционных), 2–ой родительный и 2–ой предложный ед. числа; 6 падежей мн. числа.	завод
8	8	одушевлённость, род, число, падеж	род, число, падеж	6 падежей ед. числа мужского рода 6 падежей ед. числа женского рода 6 падежей ед. числа среднего рода 6 падежей мн. числа	длинный
		род, число	род, число	мужской, женский, средний род ед. числа; мн. число	весел
	степень	степень	сильная, слабая	краснее	
	возвратность, вид, время, одушевлённость, род, число, падеж, залог	род, число, падеж	6 падежей ед. числа мужского рода 6 падежей ед. числа женского рода 6 падежей ед. числа среднего рода 6 падежей мн. числа	делающий	
16 (кр. причастия)	16 (кр. причастия)	возвратность, вид, время, род, число, залог	род, число	мужской, женский, средний род ед. числа; мн. число	сделан
		9	9	возвратность, вид, наклонение, лицо, род, число	наклонение, лицо, род, число

М-класс	С-класс	ГП	Свободные ГП	Спектр значений свободных ГП	Примеры
	19 (дее-при-частия)	возвратность, вид, время	время	ед. и мн. число повелительного наклонения; мужской, женский средний род ед. числа прошедшего наклонения; мн. число прошедшего наклонения; инфинитив настоящего времени несовершенного вида; прошедшее время	неся

Отсутствие у И-слова одной или нескольких характерных форм (например, форм единственного числа у слова «ножницы», форм повелительного наклонения у слова «видеть») не препятствует отнесению его к соответствующему М-классу (такая особенность словоизменения указывается в словарной статье).

П-класс — подмножество М-класса, представители которого при одинаковых значениях свободных ГП имеют одинаковые флексии. В модели рассматриваются 24 П-класса для субстантивных слов, 8 — для адъективных, 2 — для местоименных, 9 — для спрягаемых (см. приложение I). К первому П-классу субстантивных И-слов относятся, например, существительные «кран» и «дятел» (флексии: «-Ø», «-а», «-у», «-Ø» или «-а», «-ом», «-е» для шести традиционных падежей единственного числа; «-ы», «-ов», «-ам», «-ы» или «-ов», «-ами», «-ах» для множественного); к 4-му П-классу адъективных И-слов — прилагательное «хороший», существительное «прохожий»; ко 2-му П-классу местоименных И-слов — местоимение «мой», прилагательное «лисий», порядковое числительное «третий».

С-класс образуют слова и конструкции, имеющие общий набор ГП и общие синтаксические функции. Для большинства И-слов номер С-класса и соответствующий набор ГП совпадают с номером и набором ГП М-класса. Так, многие существительные С-класса 7 и относятся и к 7-му М-классу. Однако некоторые слова изменяются по «необычной» модели: существительные «сказуемое», «столовая» склоняются как представители 8-го М-класса, для существительного «кабельтов» характерно местоименное склонение. В подобных ситуациях в описании значения основы указывается и синтаксический класс, а иногда и значения «дополнительных» ГП (например, вида и залога для причастий — С-класс 18, склоняющихся по модели 8-го М-класса) — см. § 4.

§ 2. Исключения.

Хотя П-классы задают более детальную классификацию сочетаемости основ с флексиями, чем традиционные типы склонения ж спряжения, они недостаточны для описания многих частных особенностей русского словоизменения. Эти особенности можно было бы учесть с помощью ещё более дробной классификации, однако во избежании чрезмерного увеличения числа П-классов (сходный подход предложен и в [12] в модели они описываются как исключения. Такое решение проблемы соответствует общему принципу построения морфологической модели: большинство слов обрабатываются простыми средствами, обеспечивающими их быстрый анализ и синтез, а слова с различными нерегулярными явлениями (их относительно мало) обрабатываются с помощью специальных правил и, обычно, медленнее.

Исключение — это правило сочетания основы с «нестандартной» для данного П-класса и данной формы флексией. Например, «-а» — в форме именительного падежа множественного числа существительных вместо характерной для 1-го П-класса «-ы» («города», но «заводы»), «-Ø» вместо «-ов» — в родительном падеже множественного числа («сапог», но «заводов»). Исключением считается и наличие у некоторых существительных 2-го ро-

дительного и 2-го предложного падежей — «кусок сахару», «в шкафу», но «из сахара», «о шкафу». Всего в модели учитывается 26 исключений, которые представлены в таблице исключений. Исключениями считаются также следующие «нестандартные» ситуации, описание которых приведено в § 4: несовпадение для некоторых слов номеров М-класса и С-класса (например, слово «прохожий» изменяется как представитель 8-го М-класса, и поэтому относится к 8-му М-классу, а по синтаксическим признакам оно принадлежит к 7-му С-классу); отсутствие каких-либо форм слова (наиболее часто отсутствуют следующие формы слов: формы множественного числа — «борьба», формы единственного числа — «сани» — существительных; краткие формы — «советский», формы сравнительной степени — «городской», полные формы — «рад» — прилагательных; формы повелительного наклонения — «видеть», деепричастия — «ехать» — глаголов и некоторые другие); наличие супплетивных форм (например, формы «лучше» у слова «хороший»); наличие возвратной частицы «-ся» у некоторых прилагательных и существительных, т.е. у слов таких классов, набор ГП которых не содержит ГП «возвратность» («выдающийся», «учащийся»).

ТАБЛИЦА ИСКЛЮЧЕНИЙ

Тип исключения	Флексия	С-класс	П-классы	Одушевленность	Род	Число	Падеж	Наклонение	Примеры	
									С исключением	Без исключения
1	2	3	4	5	6	7	8	9	10	11
МИ1	А	7	1,3,4,5,8,9	1	1	2	1	–	мастер–мастера	артист–артисты
				2	1	2	1,4	–	глаз–глаза	алмаз–алмазы
МИ2	∅	7	1,3,5,12	1	1	2	2,4	–	грузин–грузин	бегун–бегунов
				2	1,2	2	2	–	глаз–глаз	алмаз–алмазов
МИ3	Я	19*	6,7	–	–	–	–	3	увезти–увезя	везти–везши
МИ4	Ь	9	1,2,3,4,6,8	–	0	1	–	5	слышать–слышь	бежать–беги
МИ5	∅	9	2,6,7	–	1	1	–	3	гибнуть–гиб	завянуть–завял
МИ6	ОВ	7	21	2	3	2	2	–	облако–облаков	свойство–свойств
МИ7	ЬТЕ	9	1,2,3,4,6	–	0	2	–	5	тратить–тратьте	шадить–шадите
МИ8	ШИ	19	2,6	–	–	–	–	3	лезть–лезши	гнуть–гнувши
МИ9	И	9	3	–	0	1	–	5	строить–строй	ловить–лови
МИ10	ИГЕ	9	3	–	0	2	–	5	стоять–стойте	ловить–ловите
МИ11	Я	7	2,6,7	1	1	2	1	–	слесарь–слесаря	житель–жители
				2	1	2	1,4	–	край–края	случай–случаи
МИ12	ЕЙ	7	12,14,16	1	2	2	2,4	–	ханжа–ханжей	кассирша–кассирш
				2	2	2	2	–	кегля–кеглей	неделя–недель
МИ13	ЕВ	7	25,28	2	3	2	2	–	платье–платьев	ущелье–ущелий
МИ14	И	7	21,23	2	3	2	1,4	–	яблоко–яблоки	облако–облака
МИ15	И	9	1	–	0	1	–	5	трепать–трепли	делать–делай
МИ16	ИТЕ	9	1	–	0	2	–	5	колебать–колеблите	делать–делайте
МИ17	У	7	1,3,4,5,9	2	1	1	8	–	лес — в лесу	лес — о лесе
МИ18	Ю	7	2,6,7	2	1	1	8	–	строй — в строю	строй — о строе
МИ19	У	7	1,3,4,5,9	2	1	1	7	–	кусок сахару	из сахара
МИ20	Ю	7	2,6,7	2	1	1	7	–	ложка дётю	из дётя
МТ21	У	7	1,3,4,5,9	2	1	1	7,8	–	ком снегу на снегу	из снега о снеге
МИ22	Ю	7	2,6,7	2	1	1	7,8	–	чашка чаю в чаю	из чая, о чае
МИ23	ОМУ	5	1	0	1,3	1	3	–	дядиному другу	отцову другу
МИ24	ОГО	5	1	1	1,3	1	2,4	–	дядиного друга	отцова друга
				2	1,3	1	2	–	дядиного пальто	отцова пальто
МИ25	Ь	6**	2	–	1	1	–	–	синий–синь	дальний–дален
МИ26	Е	5	1	0	1,3	1	6	–	о Петрове	о дядином

* С-класс 19 (М-класс 9) — деепричастия.

** С-класс 6 (М-класс 8) — краткие прилагательные.

§ 3. Чередования.

К особенностям словоизменения относятся и чередования в основе, например, «ева → ю» («мал–ева–ть» — «мал–ю–ю»), «ска → щ» («та–ска–ть» — «та–щ–у»), «е → ь» («пал–е–ц–Ø» — «пал–ь–ц–а»). В модели учтено 55 чередований, 47 из которых встречаются только в глаголах. Каждое чередование встречается лишь при определённых флексиях и в определённых М–классах и П–классах. Эта связь отражена в таблице чередований (стр. 39–44). Для И–слов с чередованиями достаточно рассматривать только один «стандартный» (встречающийся в наибольшем количестве форм слова) вариант основы, указывая тип и контекст чередования (как и номер исключения) в описании значения основы. Так, для стандартного варианта основы «пальц–» указывается, что при флексии «–Ø» предпоследняя буква основы «ь» заменяется на букву «е». Любое чередование в системе описывается парой чередующихся букв («ова → у»); номером чередования; номерам М–классов и П–классов в которых оно возможно; флексиями, при которых происходит данное чередование и местом чередующихся букв в основе. Относительно редкие чередования модели не учитываются, парадигмы таких слов задаются несколькими основами или основами и Н–словами, образующими «семейство» слова (основы «зай–», «зайд–» и «заш–» для «зайти»), Двумя основами задаются парадигмы существительных, основы которых различны в формах единственного и множественного числа («цыплёнк–» и «цыплят–» для «цыплёнок», «армянин–» и «армян–» для «армянин»), Хотя для таких существительных и можно было бы ввести дополнительные чередования (например, «Ø → ин»), этого не сделано по следующим соображениям: во–первых, нецелесообразно увеличивать общее число учитываемых чередований, поскольку их обработка (изменение основы) и поиск изменённого варианта в словаре всё же замедляет морфологический анализ слова; во–вторых, в данном случае чередованием связаны две разные грамматические группы форм слова (формы единственного и множественного числа) и задание взаимных ссылок не вызывает затруднений. Семейства вводятся и для слов с супплетивными формами («хороший» — «лучше») или уникальными наборами флексий (некоторые числительные).

ТАБЛИЦА ЧЕРЕДОВАНИЙ

Условные обозначения:

$a_1a_2\dots a_{n-1}a_n$ — основа слова, где a_k — k -я буква основы, $k = 1, 2, \dots, n$;

\Rightarrow — «тогда»;

\rightarrow — «преобразуется в»;

А — $a_{n-1} = o$ \Rightarrow
 $a_{n-2}oa_n \rightarrow a_{n-2}a_n$

Б — $a_{n-1} = e$ \Rightarrow
 $a_{n-2}ea_n \rightarrow a_{n-2}a_n$

В — $a_{n-1} = e$ \Rightarrow
 $ea_n \rightarrow йa_n$

Г — $a_{n-1} = e$ \Rightarrow
 $ea_n \rightarrow ьa_n$

Д — $a_n = e$ \Rightarrow
 $a_{n-1}e \rightarrow a_{n-1}ь$

Е	–	$a_n = и$	\Rightarrow	
		$a_{n-1}И$	\rightarrow	$a_{n-1}Ь$
И	–	a_n	\rightarrow	$a_nЕН$
К	–	a_n	\rightarrow	$a_nН$

А–К — лишь основные, часто встречающиеся в таблице типы программ выполнения правил чередования, частные программы приводятся непосредственно в таблице.

Флексия	М–класс	П–класс	Тип чередования	Программа	Примеры
∅	7	1	1	А	бугор–бугра
∅			2	Б	узел–узла
∅			3	В	заём–займа
∅			4	Г	лев–льва
Ь	7	2	1	А	огонь–огня
Ь			2	Б	ливень–ливня
Ь			4	Г	увалень–увальня
∅	7	3	1	А	кубок–кубка
∅			2	Б	орешек–орешка
∅			3	В	паёк–пайка
∅			4	Г	хорёк–хорька
∅	7	5	2	Б	немец–немца
∅			3	В	боец–бойца
∅			4	Г	палец–пальца
Й	7	6	4	Д	улей–ульи
∅	7	11	1	А	кукол–кукла
∅			2	Б	сосен–сосна
∅			3	В	каём–кайма
∅			4	Г	тюрем–тюрьма
∅	7	13	1	А	сказок–сказка
∅			2	Б	точек–точка
∅			3	В	чаек–чайка
∅			4	Г	серёг–серьга
Ь	7	12	1	А	кухонь–кухня
Ь			2	Б	петель–петля
∅	7	15	2	Б	овец–овца
∅	7	12	3	В	боен–бойня
∅			4	Г	спален–спальня
И	7	16	4	Д	статей–статья
И			6	Е	гостей–гостья
Ь, ЪЮ	7	18	1	А	любовь–любви
Ь, ЪЮ	7	19	1	А	вошь–вши
∅	7	21	1	А	окон–окно
∅			2	Б	кресел–кресло
∅			4	Г	писем–письмо
∅	7	23	2	Б	блюдец–блюдец
∅			4	Г	мылец–мыльце
Я	7	27	50	И	имя–имени
И	7	28	4	Д	питей–питьё
И			6	Е	ущелей–ущелье
и	5	2	4	Д	чей–чьего
и			6	Е	лисий–лисьего
∅	8	1	2	Б	удобен–удобна
∅			3	В	буен–буйный
∅			4	Г	волен–вольный
∅			34	К	образовац–образованць
∅, А, О, Ы			34	К	призвац–призванць
∅	8	2	2	Б	излишен–излишний

Флексия	М-класс	П-класс	Тип чередования	Программа	Примеры
∅			3	В	бескрасн–бескрайний
∅			4	Г	дален–дальный
∅	8	3	1	А	зорок–зоркий
∅			2	Б	тяжек–тяжкий
∅	8	3	3	В	стойкий–стойк
∅			4	Г	горький–горек
е			27	$a_n = ж \Rightarrow$ $a_{n-1}ж \rightarrow a_{n-1}Г$	строже–строгий
е			29	$a_n = ч \Rightarrow$ $a_{n-1}ч \rightarrow a_{n-1}К$	крепче–крепкий
е			41	$a_n = ш \Rightarrow$ $a_{n-1}ш \rightarrow$	тище–тихий
∅	8	7	1	А	смешон–смешной
∅			2	Б	чуден–чудной
∅			4	Г	хмелен–хмельной
ть, л, ла, ло, ли, в, вши	9	1	1	$a_n = 0 \Rightarrow$ $a_{n-1}0 \rightarrow a_{n-1}$	колоть–колю
й, йте			4	Д	лей–лью
ть, л, ла, ло, ли, в, вши			6	Е	бить–бью
—"–"			30	$a_{n-1}a_n = па, ба, ма$ \Rightarrow	колебать–колеблет
—"–"			31	$a_{n-2}a_{n-1}a_n = ова \Rightarrow$ $a_{n-3}ова \rightarrow$	рисовать–рисую
—"–"			32	$a_{n-2}a_{n-1}a_n = ева \Rightarrow$ $a_{n-3}ева \rightarrow$	жевать–жую
—"–"			33	$a_{n-2}a_{n-1}a_n = ева \Rightarrow$ $a_{n-3}ева \rightarrow$	малевать–малюю
—"–", я			35		даю–давать
ю, ешь, ет, ете, ют, ем			56	$a_n \rightarrow a_nв$	
ть, л, ла, ло, ли, в, вши			43	$a_n = ы \Rightarrow$ $a_{n-1}ы \rightarrow a_{n-1}о$	рыть–рою
—"–"			44	$a_n = е \Rightarrow$ $a_{n-1}е \rightarrow a_{n-1}о$	петь–пою
—"–"			51	$a_{n-1}a_n = ла \Rightarrow$ $a_{n-2}ла \rightarrow a_{n-2}ел$	стлать–стелю
—"–"	9	2	9	$a_{n-1}a_n = за \Rightarrow$ $a_{n-2}за \rightarrow a_{n-2}ж$	сказать–скажу
—"–"			12	$a_{n-1}a_n = са \Rightarrow$ $a_{n-2}са \rightarrow a_{n-2}ш$	писать–пишу
—"–"			14	$a_{n-1}a_n = та \Rightarrow$ $a_{n-2}та \rightarrow a_{n-2}ч$	прятать–прячу
—"–"			16	$a_{n-1}a_n = та \Rightarrow$ $a_{n-2}та \rightarrow a_{n-2}щ$	клеветать–клеветшу
ть, л, ла, ло, ли, в, вон	9	2	18	$a_{n-1}a_n = га \Rightarrow$ $a_{n-2}га \rightarrow a_{n-2}ж$	брызгать–брызжу
—"–"			19	$a_{n-1}a_n = ка \Rightarrow$ $a_{n-2}ка \rightarrow a_{n-2}ч$	плакать–плачу
—"–"			20	$a_{n-1}a_n = ха \Rightarrow$ $a_{n-2}ха \rightarrow a_{n-2}ш$	пахать–пашу
—"–"			21	$a_{n-2}a_{n-1}a_n = ска \Rightarrow$ $a_{n-3}ска \rightarrow a_{n-3}щ$	искать–ищу
—"–"			22	$a_{n-1}a_n = да \Rightarrow$ $a_{n-2}да \rightarrow a_{n-2}д$	глодать–гложу
—"–"			23	$a_{n-2}a_{n-1}a_n = ста \Rightarrow$ $a_{n-3}ста \rightarrow a_{n-3}щ$	хлестать–хлещу
ла, ло, ли, ши			34	$a_{n-1}a_{n-1} \rightarrow a_{n-1}н$	достигла–достигнет
ть, л, ла, ло, ли, в, вши			35	$a_n = а \Rightarrow a_{n-1}а = a_{n-1}$	ржать–ржу
—"–"			37	$a_n = у \Rightarrow a_{n-1}у = a_{n-1}$	мокнуть–мокнет
—"–"	9	3	2	$a_n = е \Rightarrow a_{n-1}е = a_{n-1}$	смотреть–смотрю
—"–"			5	$a_n = и \Rightarrow a_{n-1}и = a_{n-1}$	смолить–смолю
ю			7	$a_n = л \Rightarrow a_{n-1}л = a_{n-1}$	ловлю–ловит
ть, л, ла, ло, ли, в, вши			36	$a_n = я \Rightarrow a_{n-1}я = a_{n-1}$	стоять–стою

Флексия	М-класс	П-класс	Тип чередования	Программа	Примеры
—" —			52	$a_{n-1}a_n = na \Rightarrow$ $a_{n-2}на \rightarrow a_{n-2}OH$	гнать—гоню
—" —	9	4	5	$a_n = и \Rightarrow a_{n-1}и = a_{n-}$	множить—множу
—" —			35	$a_n = а \Rightarrow a_{n-1}а = a_{n-}$	слышать—слышу
—" —	9	5	2	$a_n = е \Rightarrow a_{n-1}е = a_{n-}$	видеть—видит
—" —			5	$a_n = и \Rightarrow a_{n-1}и = a_{n-}$	щадить—щадит
у			8	$a_nж \Rightarrow$ $a_{n-1}ж \rightarrow a_{n-1}з$	вожу—возит
у			10	$a_nж \Rightarrow$ $a_{n-1}ж \rightarrow a_{n-1}д$	бужу—будит
у			11	$a_nш \Rightarrow$ $a_{n-1}ш \rightarrow a_{n-1}с$	брошу—бросит
у			13	$a_nч \Rightarrow$ $a_{n-1}ч \rightarrow a_{n-1}т$	трачу—тратит
у			15	$a_nщ \Rightarrow$ $a_{n-1}щ \rightarrow a_{n-1}т$	запрещу—запретит
у			17	$a_nщ \Rightarrow$ $a_{n-1}щ \rightarrow a_{n-1}ст$	угощу—угостит
ть	9	6	24	$a_nс \Rightarrow$ $a_{n-1}с \rightarrow a_{n-1}д$	украсть—украду
ть, л, ла, ло, ли, в, вши			34	$a_n \rightarrow a_nн$	одеть—одену
—" —			35	$a_n = а \Rightarrow$ $a_{n-1}а \rightarrow a_{n-1}$	рвать—рву
—" —			38	$a_n \rightarrow a_nv$	жить—живу
л, ла, ло, ли, в, вши			39	$a_n \rightarrow a_nд$	прял—прядёт
ть, л, ла, ло, ли, в, вши			42	$a_{n-1}a_n = pa \Rightarrow$ $a_{n-2}pa \rightarrow a_{n-2}ep$	братъ—беру
—" —			45	$a_na \Rightarrow$ $a_{n-1}а \rightarrow a_{n-1}M$	жать—жму
—" —			46	$a_{n-1}ня \Rightarrow$ $a_{n-2}ня \rightarrow a_{n-1}M$	принять—приму
—" —			48	$a_{n-1}ня \Rightarrow$ $a_{n-2}ня \rightarrow a_{n-2}ЙM$	нанять—найму
—" —			49	$a_nя \Rightarrow$ $a_{n-1}я \rightarrow a_{n-1}IM$	обнять—обниму
—" —			53	$a_{n-1}a_n = va \Rightarrow$ $a_{n-2}va \rightarrow a_{n-2}OV$	звать—зову
—" —			54	$a_nя \Rightarrow$ $a_{n-1}я \rightarrow a_{n-1}H$	мять—мну
—" —			55	$a_na \Rightarrow$ $a_{n-1}а \rightarrow a_{n-1}H$	жать—жну
ти	9	7	24	$a_nс \Rightarrow$ $a_{n-1}с \rightarrow a_{n-1}д$	вести—веду
ти			25	$a_nс \Rightarrow$ $a_{n-1}с \rightarrow a_{n-1}т$	мести—мету
ти			26	$a_nс \Rightarrow$ $a_{n-1}с \rightarrow a_{n-1}б$	грести—гребу
л, ла, ло, ли			39	$a_n \rightarrow a_nд$	вёл—веду
—" —			40	$a_n \rightarrow a_nт$	смёл—смету
∅, ши	9	8	2	Б	жѣг—жгу
—" —			1	А	толок—толкла
ешь, ет, ем, ете			27	$a_nж \Rightarrow$ $a_{n-1}ж \rightarrow a_{n-1}т$	стрижет—стригу
ь			28	$a_nч \Rightarrow$ $a_{n-1}ч \rightarrow a_{n-1}т$	беречь—берегу
ь, ешь, ет, ем, ете			29	$a_nч \Rightarrow$ $a_{n-1}ч \rightarrow a_{n-1}к$	печёт—пеку
еть, ∅, ла, ло, ли, ев, ши	9	9	2	$a_{n-1}a_n = ep \Rightarrow$ $a_{n-2}ep \rightarrow a_{n-2}p$	тереть—тру

§ 4. Словарь.

Основная часть сведений о русской морфологии, используемой при анализе слов, представлена в словаре системы, структура которого непосредственно отражает структуру морфологической модели и учитывает особенности языка реализации. Так, имена словаря (идентификаторы ПЛЭНЕРа) соответствуют Н-словам, основам И-слов или флексиям, поля — описаниям значений этих объектов. Название поля определяет тип объекта: для Н-слов — MW, для невозвратных основ — MS, для возвратных — MR, для флексий в общем случае — MF, в случае исключений — МИ1, ..., МИ26 соответственно для каждого типа исключений. Для новых слов (Н-слов и И-слов) указывается дополнительное поле MN с признаком «Т». Для слов, допускающих возможность неоднозначного членения на основу и флексию в некоторых своих формах, вводятся дополнительные поля, в том числе поля MD, MDD, MG, MGG с признаками «Т». Слова имеют различные дополнительные поля в зависимости от того, совпадает ли сама основа (Н-слово) с какой-либо формой другого слова («дома» и «дома»); допускает ли какая-нибудь форма два различных членения на основу и флексию, соответствующие двум различным словам («колёт-е» и «кóл-ете»), совпадает ли форма некоторого слова с формой другого слова после применения правила чередования («печь» — существительное, и «печь» — глагол с основой «пек-») и некоторые другие.

При активации словаря поля становятся списками свойств [20] соответствующих идентификаторов (имён). Отметим, что такая организация словаря обеспечивает быстрый поиск имени и поля с помощью встроенных средств системы ПЛЭНЕР-БЭСМ (функция GET).

Для Н-слов словарная статья имеет наиболее простой вид:

$$MW \rightarrow (((\langle \text{номер С-класса} \rangle \langle \text{список значений ГП} \rangle) \langle \text{Л-С-зона} \rangle) \dots) .$$

На первом уровне описываются значения, отличающиеся грамматическими (синтаксическими) показателями («...» — возможность повторения). Для каждого такого варианта значений могут существовать несколько подвариантов с различными лексико-семантическими свойствами, последовательно перечисляемыми в Л-С-зоне. Например, в словарной статье Н-слова «инкогнито» описываются три варианта значения слова:

$$\begin{aligned} MW \rightarrow & (((7 \ 2 \ 3 \ 0 \ 0) = \text{пробывание под вымышленным именем} =) \\ & ((7 \ 1 \ 1 \ 0 \ 0) = \text{лицо, скрывающее своё имя} =) \\ & ((1) = \text{наречие} =)) . \end{aligned}$$

В этом и последующих примерах Л-С-зоны словарных статей заменены на текст, поясняющий значение соответствующего варианта. Реальная структура Д-С-зоны в данной работе не рассматривается. Из словарной статьи следует, что это слово может быть соответственно неодушевлённым существительным среднего рода, одушевлённым существительным мужского рода и наречием. Словарная статья основы имеет следующий вид:

$$MS (MP) \rightarrow ((\langle \text{М-С-зона} \rangle \langle \text{Л-С-зона} \rangle) \dots) .$$

Здесь также на первом уровне приводятся варианты, отличающиеся грамматическими, или морфосинтаксическими показателями, а в Л-С-зоне указываются лексико-семантические подварианты. С основой «пол-», например, может быть связано 5 вариантов: «Поля», «Поль», «пол(настил)», «пол(биологический)», «поле» и несколько подвариантов для последнего значения. М-С-зона имеет следующую структуру:

(<П-Ч><М-Ш><ИСКЛ>) .

Элемент <П-Ч> — номер Е-класса для слов без чередований или список вида

(<номер П-класса> <номер чередования> <список флексий, при которых происходит данное чередование>) . . .) ;

<М-ГП> — число, «склеенное» из номера М-класса и из связанных с основой значений ГП; факультативный элемент <ИСКЛ> — последовательность исключений, в которой указываются описанные выше исключения МИ1, ..., МИ26; номер С-класса или список, состоящий из номера С-класса и значений дополнительных ГП в случае несовпадения номеров М-класса и С-класса («прохожий»); «*» — в случае наличия возвратной частицы «-ся» у некоторых прилагательных и существительных («светящийся», «трудящийся»); список вида

$$((L')) \begin{cases} * \\ \langle \text{словоформа} \rangle \dots \\ (\langle \text{основа} \rangle \langle \text{М-С-зона} \rangle) \end{cases}$$

для супплетивных и отсутствующих форм. «{...}» означает присутствие какого-то одного из указанных вариантов. (L') — список значений ГП. Если форма (формы) слова, соответствующая указанным в (L') значениям ГП у данной основы отсутствует, то после (L') ставится «*», если форма слова уникальна — то конкретная словоформа, если для данного слова в словаре хранится несколько основ, то указываются ссылки на другие основы в виде (<основа><М-С-зона>). Например, элемент <ИСКЛ> для основы «ед-» слова «ехать» таков:

$$\langle \text{ИСКЛ} \rangle \rightarrow ((115000) * (111) * (113000) \\ (\text{ЕХА} (69123000)) (12600) \text{ЕХАТЬ}),$$

т.е. слова отсутствуют формы повелительного наклонения, деепричастия настоящего времени; при образовании форм прошедшего времени используется словарная статья основы «еха-»; в инфинитиве употребляется словоформа «ехать». А для одной из основ слова «цыплёнок» — «цыплят-» элемент <ИСКЛ> выглядит так:

$$\langle \text{ИСКЛ} \rangle \rightarrow ((71120) ((1110) (\text{ЦИПЛЕНК}((31(()) 71110))))),$$

где первый элемент списка указывает принадлежность слова к мужскому роду (хотя оно склоняется по образцу существительных среднего рода 21 П-класса), а второй обозначает ссылку на соответствующий вариант основы форм единственного числа.

М-С-зона для основы «пальц-» такова: «((5(4()))72100)», где 72100 — 7-й М-класс, неодушевлённость, мужской род, число и падеж любые.

Описание значений ГП некоторых основ могут находиться и в поле MS и в поле MR. Например, словарная статья основы «кол-» слов «колоть», «колоться», «Коля» и «кол» имеет такой вид:

$$(\text{MS} (((11(\text{ТЬ Л ЛА ЛО ЛИ В ВШИ})) 9120000)=\text{КОЛОТЬ=}) \\ ((1271200(70100))=\text{МАЛ ПО ИМ: КОЛЯ=}) \\ ((172100\text{МИ17})=\text{КОЛ=}) \\ \text{MR}(((11(\text{ТЬ Л ЛА ЛО ЛИ В ВШИ})) 9220000)=\text{КОЛОТЬСЯ=}) \text{MDD T} .$$

Словарная статья упоминавшейся выше основы «пол-» имеет следующий вид:

(MS (((26 72300)=ПОЛЕ/ПОЛЕ2=
 (((12 71200)=ДЕВ-ПО-ИМ:ПОЛЯ=
 (2 71100)=МАЛ-ПО-ИМ:ПОЛЬ=
 (1 72100 МИ17)=ПОЛ/НАСТИЛ/=)
 (1 72100)=ПОЛ/ВИОЛ/=)))

Описание флексии имеет следующую структуру: <вид поля> → ((<номер М-класса>(<список номеров П-классов> <список определяемых флексией значений свободных ГИ>)...)). Вид поля — MF (в общем случае) или МИ1, ..., МИ26 (для исключений). Описание одной флексии также может находиться в нескольких различных долях. Полное описание рассматриваемого в модели множества флексий русского языка в форме, используемой при морфологическом анализе, приводится на стр. 25. Описание флексий, чередований и исключений для процедур синтеза см. в приложении 3 (стр. 75).

Имена могут иметь и другие поля различного назначения. Так, жена для слов, появление которых в тексте «предсказывает» появление зависимых от них слов и групп (для языковых предикатов и идиом), имеют поля SL, по которым выбираются соответствующие подварианты значений и иницируется поиск предсказываемых элементов [10].

ТАБЛИЦА ФЛЕКСИЙ

1 Флексия	2 Поле	3 М-класс	4 П-классы	5 Значения * ГП	
0	MF	7	1,3,4,5,8,9	011(1 4)	
			8	112(2 4)	
			21,23,27	032(2 4)	
			11,13,14,15	022(2 4)	
		5	1	011(1 4)	
		8	1,2,3,4,5,7,8	11	
		9	8,9	003011	
		МИ2	7	1,3,5,12	0(1 2)2(2 4)
				1,3,5	011(1 4)
		МИ5	9	2,6,7	003011
A	MF	7	1,3,4,5,8,9	011(2 4)	
			11,13,14,15	0211	
			21,23	0312	
			21,23,27	032(1 4)	
		8	1,3,4,5,7,8	21	
		5	1	0211	
			1	0(1 3)1(2 4)	
		9	2,4	021	
		Ш1	7	1,3,4,5,8,9	012(1 4)
				1,3,4,5,8,9	011(2 4)
B	MF	9	1,2,3,4,5,6	103	
E	MF	7	1,2,3,4,5,6,8,9	0116	
			8	1121	
			11,12,13,14,15,16	021(3 6)	
			23,25,26,28	031(1 4)	
			21,23,26,28	0316	
		8	2,4,5,6	31	
			3,8	2	
		5	2	031(1 4)	
		МИ26	5	1	0(1 3)16

* Значения ГП для различных М-классов даны в следующем порядке: Класс 5 — од., род, число, падеж; М-класс 7 — од., род, число, падеж; М-класс 8 — од., род, число, падеж (либо сравнительная степень, либо род и число для кратких форм); М-класс 9 — возвр., вид, од., лицо, род, число (либо возвр., вид, время для деепричастий)

1 Флексия	2 Поле	3 М–класс	4 П–классы	5 Значения * ГП	
И	MF	7	2,3,4,6,7,9	012(1 4)	
			12,13,14,16	0212	
			12,13,14,16,17,18,19	022(1 4)	
			17,18,19	021(2 3 6)	
			25	2316	
			27	231(2 3 6)	
		5	2	002(1 4)	
		8	2,3,4,6,8	02	
		9	2,3,4,5,6,7,8,9	005201	
		МИ14	7	21,23	032(1 4)
МИ15	9	1	005201		
Й	MF	7	6,7	011(1 4)	
			16,17	022(2 4)	
			25,28	2322	
		5	2	011(1 4)	
		8	6	11	
		9	1	005201	
		МИ9	9	3	005201
Л	MF	9	1,2,3,4,5,6,7	003011	
О	MF	5	1	031(1 4)	
		7	21,27	031(1 4)	
		8	1,3,7,8	31	
У	MF	7	1,3,4,5,8,9	0113	
			11,13,14,15	0214	
			21,23	0313	
		5	1	0214	
			1	0(1 3)13	
		9	2,4,5,6,7,8,9	021101	
			2,4,5,6,7,8,9	012101	
		МИ17	7	1,3,4,5,9	211(3 8)
		МИ19	7	1,3,4,5,9	211(3 7)
		МИ21	7	1,3,4,5,9	211(3 7 6)
Ы	MF	5	1	002(1 4)	
		7	1,5	012(1 4)	
			11,15	0212	
			11,15	022(1 4)	
		8	1,5,7	02	
Ь	MF	7	2	011(1 4)	
			12	022(2 4)	
			18,19	021(1 4)	
			9	8	00(2 6)000
		МИ5	9	1,2,3,4,6	005201
Ю	MF	7	2,6,7	0113	
			12,16,17	0214	
			25,26,28	2313	
		5	2	0214	
Ю	MF	9	1,3	021101	
			1,3	012101	
		МИ18	7	2,6,7	211(3 8)
		МИ20	7	2,6,7	211(3 7)
		МИ22	7	2,6,7	211(3 7 8)
Я	MF	7	2,6,7	011(2 4)	
			12,16,17	0211	
			25,26,28	2312	
			25,26,28	232(1 4)	
			27	231(1 4)	
			5	2	0211
		8	2,6	21	
		9	1,3,5,6,7	021	
		МИ3	9	6,7	013
				6,7	021

Флексия	Поле	М–класс	П–классы	Значения * ГП
1	2	3	4	5
	МИ11	7	2,6,7	012(1 4)
			2,6,7	011(2 4)
АМ	MF	7	1,3,4,5,8,9	0123
			11,13,14,15,19	0223
			21,23,27	0323
АТ	MF	9	4	021302
			4	012302
АХ	MF	7	1,3,4,5,8,9	0126
			11,13,14,15,19	0226
			21,23,27	0326
АЯ	MF	8	1,3,4,5,7,8	0211
ЕВ	MF	7	5,6,7	012(2 4)
	МИ13	7	25,28	032(2 4)
ЕЕ	MF	8	2,4,5,6	031(1 4)
			1,2,4,5,6,7	2
ЕЙ	MF	7	2,4	012(2 4)
			12,14,15,16,17	0215
			18,19	022 (2 4)
			26	2322
	5	2	021(2 356)	
	8	2,4,5,6	021(2 356)	
	МИ12	7	12,14,16	022(2 4)
ЕМ	MF	5	2	0(1 3)16
		7	2,4,5,6,7	0115
			23,25,26,27,28	0315
			8	2,4,5,6
		9	1,2,6,7,8,9	021102
	1,2,6,7,8,9	012102		
ЕТ	MF	9	1,2,6,7,8,9	021301
			1,2,6,7,8,9	012301
ЕЮ	MF	5	2	0215
		7	12,14,15,16,17	0215
		8	1,3,7,8	0215
ИЕ	MF	8	2,3,4,6,8	002(1 4)
ИЙ	MF	8	2,3,4,6	011(1 4)
ИМ	MF	5	2	0(1 3)15
			2	0023
		8	2,3,4,6,8	0(1 3)15
			2,3,4,6,8	0023
			3,4,5	021102
	3,4,5	012102		
ИТ	MF	9	3,4,5	021301
			3,4,5	012301
ИХ	MF	5	2	002(2 4 6)
		8	2,3,4,6,8,9	002(2 4 6)
ЛА	MF	9	1,2,3,4,5,6,7,8,9	003021
ЛО	MF	9	1,2,3,4,5,6,7,8,9	003031
ЛИ	MF	9	1,2,3,4,5,6,7,8,9	003002
ОВ	MF	7	1,3	012(2 4)
	МИ6	7	21	032(2 4)
ОЕ	MF	8	1,3,7,8	031(1 4)
ОЙ	MF	5	1	021(2 356)
		7	11,13	0215
		8	7,8	011(1 4)
			1,3,7,8	021(2 356)
ОМ	MF	5	1	0(3 1)16
		7	1,3,8,9	0115
			21	0315
		8	1,3,7,8	0(1 3)16
ОЮ	MF	5	1	0215
		7	11,13	0215

Флексия	Поле	М–класс	П–классы	Значения * ГП
1	2	3	4	5
		8	1,3,7,8	0215
ТЙ	MF	9	7	00(2 6)000
ТЬ	MF	9	1,2,3,4,5,6	00(2 6)000
УТ	MF	9	2,6,7,8,9	021302
			2,6,7,8,9	012302
УЮ	MF	8	1,3,4,5,7,8	0214
ШИ	MF	9	7,8,9	003
	МИ8	9	2,6	003
ЬЕ	MF	8	1,5,7	002(1 4)
ЬЙ	MF	8	1,5	001(1 4)
ЬМ	MF	5	1	0(1 3)15
			1	0023
		8	1,5,7	0(1 3)15
			1,5,7	0023
ЬХ	MF	5	1	002(2 4 6)
		8	1,5,7	002(2 4 6)
ЬЮ	MF	7	18,19	0215
ЮТ	MF	9	1	021302
			1	012302
ЮЮ	MF	8	2,6	0214
ЯМ	MF	7	2,6,7	0123
			12,16,17,18	0223
			25,26,28	2323
ЯИ	MF	9	3,5	021302
			3,5	012302
ЯХ	MF	7	2,6,7	0126
			12,16,17,18	0226
			25,26,28	2326
ЯЯ	MF	8	2,6	0211
АМИ	MF	7	1,3,4,5,8,9	0125
			11,13,14,15,19	0225
			21,23,27	0325
ВШИ	MF	9	1,2,3,4,5,6	003
ЕГО	MF	5	2	0(1 3)1(2 4)
		8	2,4,5,6	0(1 3)1(2 4)
ЕТЬ	MF	9	9	00(2 6)000
ЕМУ	MF	5	2	0(1 3)13
		8	2,4,5,6	0(1 3)13
ЕТЕ	MF	9	1.2.6,7,8,9	021202
			1,2,6,7,8,9	012202
ЕШЬ	MF	9	1,2,6,7,8,9	021201
			1,2,6,7,8,9	012201
ИМИ	MF	5	2	0025
		8	2,3,4,6,8	0025
ИТЕ	MF	9	2,3,4,5,6,7,8,9	005202
			3,4,5	021202
			3,4,5	012202
	МИ16	9	1	005202
ИШЬ	MF	9	3,4,5	021202
			3,4,5	012202
ЙТЕ	MF	9	1	005202
	МИ10	9	3	005202
ОГО	MF	8	1,3,7,8	0(1 3)1(2 4)
	МИ24	5	1	0(1 3)1(2 4)
ОМУ	MF	8	1,3,7,8	0(1 3)13
	МИ23	5	1	0(1 3)13
ЬМИ	MF	5	1	0025
		8	1,5,7	0025
ЬТЕ	МИ7	9	1,2,3,4,6	005202
ЯМИ	MF	7	2,6,7	0125
			12,16,17,18	0225

Флексия	Поле	М–класс	П–классы	Значения * ГП
1	2	3	4	5
			25,26,28	2325

ТАБЛИЦА ФЛЕКСИЙ, ИСПОЛЬЗУЕМАЯ ПРИ ИСПРАВЛЕНИИ ОШИБОК, СВЯЗАННЫХ С НЕПРИМЕНЕНИЕМ НУЖНОГО ПРАВИЛА ИСКЛЮЧЕНИЯ ПРИ СЛОВОИЗМЕНЕНИИ

Флексия	Флексии–исключения
∅	МИ6 ОВ МИ12 ЕЙ МИ13 ЕВ МИ25 Ь
А	МИ14 И МИ24 ОГО
В	МИ8 ШИ
И	МИ1 А МИ4 Ь МИ9 И МИ11 Я
Й	МИ12 ЕЙ МИ13 ЕВ МИ15 И
Л	МИ5 ∅
У	МИ23 ОМУ
Ы	МИ1 А
Ь	МИ2 ∅ МИ12 ЕЙ
Я	МИ14 И
ВШИ	МИ8 ШИ
ЕВ	МИ2 ∅
ЕЙ	МИ6 ОВ
ИТЕ	МИ7 ЬТЕ МИ10 ИТЕ
ЙТЕ	МИ7 ЙТЕ МИ16 ИТЕ
ОВ	МИ2 ∅
ОМ	МИ26 Е
ШИ	МИ Я

§ 5. Средства сопровождения словаря системы.

К процедурам сопровождения словаря СИМТЕМЫ относятся процедуры КОД, 333, КСЛ, ДЕМ, ХХХ, AMID1 и HELP.

Процедуры автоматизированного пополнения словаря системы по данным из словаря А.А. Зализняка [12] — КОД и 333 — описаны в главе 3, §1.

Процедура КСЛ на вход получает один из словарей системы TULIPS–2. Результатом её работы является наглядная распечатка словаря, где основы слов и Н–слова располагаются в алфавитном порядке. Для каждого основы и Н–слова указывается его номер в обрабатываемом словаре (порядковый номер словарной статьи во внешнем представлении словаря — в долговременной памяти (системы), в каких полях представлена словарная информация об этих объектах, и количество различных интерпретаций объекта для каждого поля. Например, если обрабатывается словарь, состоящий из двух основ: «пол–» (которой соответствуют слова «поле», «Поля», «Поль» и два значения слова «пол») и «кол–» (которой соответствуют слова: «кол», «Коля», «колоть», «колоться», и есть специальное поле MDD, свидетельствующее о совпадении ФОРМЫ глагола «колет» с существительным «колет»), то результат работы будет такой:

ОСНОВА	№	MS	MR	MW	СПЕЦПОЛЯ
КОЛ	2	MS 3	MR 1		MDD
ПОЛ	1	MS5			

Процедура КСЛ предназначена для контроля текущего содержимого словаря. Есть и другие вспомогательные процедуры, в том числе: процедуры, распечатывающие по алфавиту только слова определённого М-класса, только Н-слова (поле MW), или основы, описанные в поле MR; процедуры, подсчитывающие общее число слов в словаре и число слов, принадлежащих каждому М-классу, и некоторые другие.

Процедура ДЕМ — функция от четырёх аргументов: первый — снова слова из словаря, второй — «В» или «НВ» (соответствующий полю MR или MS), третий — номер интерпретации данной основы в словаре системы, четвёртый — либо «ВСЕ», когда надо получить все формы данного слова, либо набор значений свободных ГП для какой-то одной словоформы; одушевлённость, род, число, падеж обозначаются первыми буквами их конкретного названия «О», «Ж», «Р». «Е», например); наклонение — как «НАСТ». «БУД», «ПРОШ», «СОСЛ», «ПСВ» и «ИНФ»; сравнительная степень — «СИЛ» (сильная) или «СЛ» (слабая). Для синтеза конкретных словоформ происходит обращение к процедуре MORS (см. § 7, стр. 34). Примеры работы процедуры ДЕМ:

```
(КОН НВ 2 ВСЕ) ДЕМ→
      КОНЬ КОНЯ КОНЮ КОНЯ КОНЕМ КОНЕ
      КОНИ КОНЕЙ КОНЯМ КОНЕЙ КОНЯМИ КОНЯХ
```

```
(КОЛ В 1 (ПРОШ 1 Ж Е)) ДЕМ→
      КОЛОЛАСЬ
```

Процедура ХХХ в режиме диалога с пользователем строит все несовпадающие по написанию формы слова. Пользователь при работе с этой процедурой должен набирать на клавиатуре терминала слова и соответствующие им словарные статьи из словаря А. А. Зализняка. Получив такие данные, ХХХ обращается к процедуре КОД, которая автоматически преобразует их в словарную статью системы TULIPS-2. Из полученной словарной информации ХХХ выделяет номера М-класса и П-класса, определяет наличие чередований и исключений (всех типов). С учётом этих данных она формирует и печатает словоформы, применяя нужные правила чередования и проверяя, не совпадает ли по написанию каждая последующая словоформа с какой-нибудь из предыдущих. Примеры работы процедуры ХХХ:

```
(ЗЕМЛЯ (Ж 2 * А)) ХХХ→
      ЗЕМЛЯ ЗЕМЛИ ЗЕМЛЕ ЗЕМЛЮ ЗЕМЛЕЙ
      ЗЕМЕЛЬ ЗЕМЛЯМ ЗЕМЛЯМИ ЗЕМЛЯХ
```

```
(ГОНЧАЯ (ЖО (П 4 А))) ХХХ→
      ГОНЧАЯ ГОНЧЕЙ ГОНЧЮ ГОНЧИЕ ГОНЧИХ ГОНЧИМ
      ГОНЧИМИ
```

Процедуры ДЕМ и ХХХ являются средством проверки правильности морфологической модели и словарных статей системы, в первую очередь словарных статей, сформированных вручную. Написано также несколько вариантов процедур ДЕМ и ХХХ, которые строят все формы известного слова по любой его словоформе (в случае неоднозначной интерпретации — все возможные парадигмы, соответствующие данной словоформе) с учетом ошибок, связанных с несвоевременным переключением регистра клавиатуры терминала и случайными удвоениями букв; а также процедуры, строящие все формы для всех возможных интерпретаций незнакомого слова по любой его словоформе.

Процедуры AMID1 и HELP — процедуры уточнения содержимого словаря в конце сеанса работы системы (автоматического и с помощью пользователя) — описаны в главе 3, § 3.

§ 6. Процедуры морфологического уровня. Анализ.

Морфологический анализ словоформы проводят процедуры WANL, MORF, STFL, GRAD и MORN. Процедуры WANL, MORF и STFL написаны М.Г. Мальковским.

Ведущая процедура морфологического уровня WANL получает от объемлющих процедур одну из словоформ входного текста — w , предсказания — описания её ожидаемых характеристик, и номер словоформы предложение, WANL с учётом предсказаний и установленного объемлющими блоками режима (выдать все варианты, только предсказанные и другие) управляет анализом w и отбирает трактовки, включаемые в окончательный результат. Если список предсказаний непуст, то WANL после обращения к процедуре MORF, а в случае незнакомых слов и к MORN, разбивает результат их работы на две группы: предсказанные и непредсказанные.

Если какую-либо словоформу нужно проанализировать повторно (например, с новыми предсказаниями), обращение к процедуре WANL не происходит. Многовариантность морфологического анализа обеспечивает внешняя по отношению к WANL процедура PRW.

Первоначально в системе перебор различных вариантов осуществляется с использованием встроенных средств языка ПЛЭНЕР-БЭСМ — механизма возвратов. В частности, в процедуре MORF при выборе одного из вариантов разбиения на основу и флексию ставилась «именованная развилка» [[20], стр. 45-46]. Если очередной вариант разбиения на последующих этапах анализа оказывался неудовлетворительным, вырабатывался «неуспех», возвращавший программу к этой развилке. Однако такой режим оказался неэкономным (с точки зрения машинного времени), поскольку он предусматривал выполнение обратных операторов (для восстановления первоначального состояния программ), а динамически точка, выработавшая «неуспех», могла оказаться очень «далеко» от развилки. Поэтому в настоящее время перебор вариантов реализован таким образом: все возможные варианты получаются сразу (после единственного обращения к процедурам WANL и MORF), часть из них обрабатывается полностью и передаётся объемлющим процедурам синтаксического анализа, а другая часть — запоминается в частично обработанном виде.

Для знакомых слов возможность неоднозначного разбиения на основу и флексию определяется по специальным признакам (например, специальному полю MDD у основы). Для таких слов, даже если первоначальное разбиение оказалось приемлемым (процедура STFL установила, что основа сочетается с флексией), ищется и другое разбиение с более короткой флексией. Для незнакомых, слов строятся все возможные варианты разбиения слова, и для каждой флексии последовательно учитывается вся связанная с ней информация из поля MF и из полей МИ1, ..., МИ26.

Возможны следующие варианты результатов работы процедуры WANL:

- 1) если слово известное и анализируется с предсказаниями, то $(() W_p W_n (PR () DF))$;
- 2) если слово известное и анализируется без предсказаний, то $(W_l () () () () ())$;
- 3) если слово незнакомое, анализируется без предсказаний, и обращение к процедуре MORN было только один раз, то $(W_l () () () 1 RR)$;
- 4) если слово незнакомое, анализируется с предсказаниями и есть предсказанный результат, то $(() W_p W_n (PR \cup RR))$;
- 5) если слово незнакомое, анализируется с предсказаниями и предсказанного результата нет, то $(() () W_n (PR \cup RR))$;

- б) если слово незнакомое, анализируется с предсказаниями, предсказанного результата нет и вся информация о слове уже обрабатывалась процедурой MORN то $((() W_n, (PR \cup ()))$; где PR — предсказание, W_p — результат работы процедуры MORF или MORN, удовлетворяющий предсказанию; W_n — результат, не удовлетворяющий предсказанию; W_l — результат работы процедуры MORF или MORN в случае анализа без предсказаний; DF — 0, если есть результат, не удовлетворяющий предсказаниям, и () — если нет; RR — информация, оставшаяся необработанной после V обращений к процедуре MORN.

Непосредственно анализирует w процедура MORF (функция от одного аргумента w). Сначала она ищет в словаре имя w с полем MN, т.е. пытается воспринять w как неизменяемое. Если поиск неуспешен, то процедура пытается отделить постфикс, а затем строит все возможные (с учётом длины словоформы) разбиения словоформы на основу и флексию, последовательно рассматривая w как основу с флексией «-∅», а затем как основу с флексиями длиной 3, 2, 1. Входная словоформа w расщепляется и её части ищутся в словаре как имена с полями MS (MR) и MF. При успехе на одном из этапов анализ прекращается, если в словаре не указана возможность неоднозначного членения (т.е. с основой не связаны специальные поля MD, MDD, MG или MGG).

Проверку правильности членения (сочетаемости основы с флексией) и определение по флексии значений свободных ГП осуществляет процедура STFL. STFL — функция от четырёх аргументов: словарной информации об основе слова (содержимое полей MS, MR); информации о флексии (содержимое поля MF); основы и флексии. Так, если, например, системе встретилось незнакомое слово «жарит», а в словаре есть основа слова «жара» — «жар-», то MORF, отщепив флексию «-ит», найдёт в словаре и основу «жар-» и флексию «-ит». Но в результате проверок процедуры STFL (не совпадает номер M-класса основы из словаря и номер M-класса, при котором возможна флексия «-ит») слово «жарит» будет воспринято как незнакомое системе.

С использованием информации метауровня STFL может также «игнорировать» некоторые ошибки на стыке основы (знакомого слова) и флексии. Если вариант, противоречащий словарной информации «понят», работа анализатора не прерывается, хотя пользователю посылается сообщение об ошибке. Например, если системе встретилась словоформа «стригёт», то, «поняв» ошибочный вариант основы, STFL сообщает пользователю: «ДОПУЩЕНА ОШИБКА В СЛО-ВОИЗМЕНЕНИИ. ПРИ ДАННОЙ ФЛЕКСИИ «-ЕТ» В ОСНОВЕ ПРОИСХОДИТ ЧЕРЕДОВАНИЕ Г → Ж». А при анализе ошибочной словоформы «глазов» пользователю сообщается, что в данной форме слово имеет нестандартную флексию «-∅». Процедура STFL обнаруживает и ошибки, связанные с употреблением форм, которые у данного слова отсутствуют. Например, форм множественного числа у существительного «борьба», кратких форм и сравнительной степени у прилагательного «городской». В подобных случаях STFL может обратиться за разъяснениями к пользователю. Отметим, что в соответствии с данными метауровня ошибки в слове обнаруживаются только в тех случаях, когда флексия соответствует M-классу и П-классу основы, записанной в словаре (например, «-ов» для слова «глазов»), если же флексия невозможна в данном M-классе или П-классе (например, «-ит» для существительного «жара»), то, в зависимости от установленного режима, либо слово воспринимается как незнакомое, либо система обращается за разъяснениями к пользователю. Процедура STFL уточняет некоторые «связанные» значения ГП, например, падеж по одушевлённости («стола» — только родительный падеж, «кота» — родительный или винительный падеж). Результатом работы процедуры STFL является список

$((((S L) \beta, \beta) \dots))$,

где S — C -класс, L — набор значений ГП, β — L - C -зона, β_r — информация для синтеза, которая извлекается из морфологической зоны словарной статьи и имеет следующий вид: ($=R$ <информация для синтеза>). Информация для синтеза — либо H -слово, либо морфологическая зона словарной статьи (например, (ПН ((2 2(Б)) 72100)) для слова «пень»), либо ссылка на соответствующую морфологическую информацию в случае соответствия одному слову «семейства» основ («цыпленк-» и «цыплят-» для слова «цыпленок»),

Когда MORF, отщепив окончание, не может найти получившуюся основу в словаре, происходит обращение к процедуре GRAD), которая пытается применить к основе правила чередования.

Процедура GRAD — функция от двух аргументов. Первый аргумент представляет собой список, состоящий из двух элементов: числа, обозначающего длину флексии и самой флексии. Второй аргумент — основа анализируемого слова. При работе процедуры в зависимости от длины флексии (0, 1, 2 или 3) происходит переход на метки M_0 , M_1 , M_2 или M_3 , где в соответствии с конкретным окончанием и (или) определёнными свойствами основы (оканчивается на определённую букву, имеет определённую предпоследнюю букву и т.п.) последовательно применяются возможные правила чередования. В большинстве случаев (когда в системе учитываются 55 чередований) для каждого слова просматриваются лишь 1–2 правила чередования, в редких случаях — до восьми. Поэтому процедуры обработки чередований не сильно замедляют прогресс анализа слов. Например, при анализе слова «брошу» по флексии «-у» процедура сразу сокращает число возможных чередований до 6, а по последней букве основы «ш» определяет и применяет единственное в данном случае правило чередования («ш → с»).

Применив правило чередования, GRAD ищет получившуюся основу в словаре, и, если она там записана, проверяет, используя словарную информацию, может ли у данной основы при данной флексии быть предполагаемое чередование. Когда данное чередование у полученной основы невозможно или её не оказалось в словаре. GRAD пробует применить другое чередование, если ещё возможно. Если все применяемые GRAD правила чередования оказались неподходящими, то результатом работы процедуры является пустой список и дальнейший анализ слова проводит процедура обработки незнакомых слов MORN, описанная в главе 3. В случае успешного применения правила чередования, результатом работы процедуры GRAD является список, состоящий из полученной основы в круглых скобках и соответствующей ей словарной информации. Например, при анализе слова «пламя» процедура GRAD применяет чередование 50 («∅ → ен»), приписав к концу основы две буквы «ен»:

[GRAD (1 Я) ПЛАМ] → ((ПЛАМЕН) (((29(50(Я)) 72300)=ПЛАМЕН=))).

Найдя основу «пламен-» в словаре и убедившись, что у неё есть чередование 50 при флексии «-я» процедура GRAD выдала указанный результат.

Если в словаре есть основа слова «кружка» — «кружк-», и системе впервые встретилось слово «кружок», то при обращении к процедуре GRAD будет применено чередование 1 («о → ∅») и полученная основа будет найдена в словаре. Но у основы «кружк-» есть только чередование 2 («е → ∅»), поэтому слово «кружок» воспримется как незнакомое системе. Процедура GRAD в этом случае выдаёт пустой список:

[GRAD (0 =) КРУЖОК] → ().

где « \Rightarrow » обозначает пустую флексию.

Если же процедуре MORF не удаётся воспринять w как форму знакового слова, то результатом её работы является список всех возможных членений w на основу и флексию и всех возможных значений ГП для каждой флексии, который является входной информацией для процедуры обработки незнакомых слов MORN (см. главу 3, § 2). Пока (на этом этапе анализа) не учитываются морфологические свойства основы (их учитывает процедура MORN).

Если же слово оказалось знакомым системе, то результатом работы процедуры MORF является результат работы процедуры STFL. Например, при анализе словоформы «коне» MORF выдаёт следующий результат:

```
(( (7 211 6) (=R(КОН(1 72100 МИ17)) )=КОН=)
  ((7 111 6) (=R(КОН(2 7ПСО))=КОНЬ=)),
```

где 1-ый элемент списка соответствует слову «кон», стоящему в форме единственного числа предложного падежа, а 2-ой — слову «конь» в той же форме.

Отметим, что морфологический анализатор система работает достаточно быстро: полная обработка словоформы с фильтрацией по семантическим предсказаниям и подготовкой информации к запоминанию требует от 0.01 до 0.20 сек.

§ 7. Процедуры морфологического уровня. Синтез.

Знания системы, используемые при синтезе словоформ (см. приложение 3), структурированы таким образом, чтобы обеспечить наиболее быстрое «вычисление» флексии, приписываемой к рассматриваемой основе. Данные о флексиях для каждого С-класса, о чередованиях и исключениях представлены в виде значений соответствующих констант. Действия, которые необходимо произвести, чтобы применить k -е чередование, описаны в k -ом элементе значения (списка) константы чередований. Флексия, которую надо приписать основе, принадлежащей, например, 7-му С-классу 3-му П-гласу для получения соответствующего слова в форме родительного (2-го) падежа единственного (1-го) числа, находится в константе 7-го С-класса в 3-ем элементе её значения на первом уровне, в 1-ом элементе второго уровня на 2-ом месте.

Синтезирует конкретные словоформы процедура MORS. MORS — функция от двух аргументов: информации для синтеза (см. § 6, стр. 31) — в общем случае это основа и морфологическая зона её словарной статьи, и списка ($S L$), где указаны значения ГП синтезируемой формы слова. Процедура MORS из информации для синтеза выделяет номер П-класса и данные о наличии чередований и исключений, а из списка ($S L$) — номер С-класса и значения свободных ГП. используя эти данные, процедура выполняет необходимые чередования букв в основе, выбирает нужную флексию и «приклеивает» её к основе. Результатом работы процедуры MORS является либо нужная форма слова, либо основа, к концу которой приписан символ «*», если требуемой формы слова нет (например, сравнительной степени у причастия), при этом пользователь получает соответствующее сообщение, «*» «приписывается» к основе также в том случае, когда значения некоторых ГП неопределены. Примеры работы процедуры MORS:

```
(( (ПН((2 2(Ь)))71200)) (7 2 1 1 1)) MORS → ПЕПЬ
```

((РАД(1 80000((0 0 0 0)*))(8 0 1 1 1)) MORS→ РАД* — «ФОРМЫ МУЖСКОГО РОДА ЕДИНСТВЕННОГО ЧИСЛА ИМЕНИТЕЛЬНОГО ПАДЕЖА У ДАННОГО ЗНАЧЕНИЯ СЛОВА С ОСНОВОЙ — РАД НЕТ»

((ПН((2 2(Ь)))71200))(7 2 1 1 0) MORS→ ПЕН* — «ЗНАЧЕНИЕ ГП «ПАДЕЖ» НЕОПРЕДЕЛЁННО.

Среднее время синтеза одной словоформы — 0.04 сек.

Процедура МОРС по основе и морфологической зоне её словарной статьи строит все формы данного слова, обращаясь к процедуре MORS. МОРС формирует список флексий (с учётом всех ТИПОВ исключений) и в цикле «приклеивает» флексии, применяя при этом нужные правила чередования. Примеры работы процедуры МОРС

(ЛЕС(1 72100 МИ1 МИ19)) MORS→

ЛЕС	ЛЕСА
ЛЕСА	ЛЕСОВ
ЛЕСУ	ЛЕСАМИ
ЛЕС	ЛЕСА
ЛЕСОМ	ЛЕСАМИ
ЛЕСЕ/ЛЕСУ	ЛЕСАХ

(ГОРОДСК (3 80000 ((0 0)*(0)*)) MORS→

ГОРОДСКОЙ	ГОРОДСКАЯ	ГОРОДСКОЕ
ГОРОДСКОГО	ГОРОДСКУЮ	ГОРОДСКОГО
ГОРОДСКОМУ	ГОРОДСКОЙ	ГОРОДСКОМУ
ГОРОДСКОЙ/ГОРОДСКОГО	ГОРОДСКУЮ	ГОРОДСКОЕ
ГОРОДСКИМ	ГОРОДСКОЙ	ГОРОДСКИМ
ГОРОДСКОМ	ГОРОДСКОЙ	ГОРОДСКОМ
ГОРОДСКИЕ	ГОРОДСК*	
ГОРОДСКИХ	ГОРОДСК*	
ГОРОДСКИМ	ГОРОДСК*	
ГОРОДСКИЕ/ГОРОДСКИХ	ГОРОДСК*	
ГОРОДСКИМИ	ГОРОДСК*	
ГОРОДСКИХ		

Глава 3. АДАПТАЦИЯ И ОБУЧЕНИЕ СИСТЕМЫ НА ЛЕКСИЧЕСКОМ И МОРФОЛОГИЧЕСКОМ УРОВНЯХ

Адаптация и обучение на лексическом уровне предполагают формирование и запоминание в долговременной памяти системы словарных статей для незнакомых слов; на морфологическом уровне — изменение некоторых морфологических показателей известных системе и новых слов (см. § 4, стр. 52).

Отметим, что аппарат адаптации и обучения разработан на базе полной и подробной морфологической модели, описанной во второй главе. И, таким образом, общность и достоверность полученных результатов гарантируется опробованием их на реальном и обширном фактическом материале. Важно также, что предлагаемые в работе методы и алгоритмы взаимосвязанно используются в целостной экспериментальной системе. Эти методы предполагают комплексное использование различных источников информации о незнакомом языковом объекте: структуры самого объекта, контекста, метауровня языковых знаний системы, указаний пользователя.

§ 1. Автоматизированное пополнение базы знаний системы о языке.

Работа средств адаптации и обучения проходит долго и утомительно для пользователя, если почти каждое слово, встретившееся при диалоге, незнакомо системе. Неизбежная неопределённость синтаксических предсказаний, невозможность перехода на семантический уровень анализа приводят к тому, что системе приходится слишком часто обращаться за разъяснениями к пользователю. Поэтому, в полном соответствии с принципами необременительного диалога и выделения базовой совокупности ЯЗС (см. гл. 1, § 1, стр. 8, § 2, стр. 9), для системы TULIPS-2 разработаны и реализованы средства автоматизированного формирования базового словаря. Эти средства используются, например, при формировании базового лексикона новой проблемной области: большие массивы словарной информации вводятся в специальном режиме «пополнения» с помощью сервисных процедур формирования словаря системы. Составляются такие массивы по результатам обработки характерных текстов или опроса потенциальных пользователей. Оба указанных способа были проверены на практике: при формировании базового лексикона для проблемной области «Арифметические задачи» просматривались тексты условий задач из школьных учебников по математике (для 1–2 классов средней школы), а при формировании лексикона для проблемной области «ЛУЧ» (ЛИСП-УЧИТЕЛЬ) студенты, отвечавшие за формирование словаря, опрашивали начавших изучать ЛИСП третьекурсников кафедры. Их вопросы о ЛИСПе протоколировались и в дальнейшем обрабатывались (так же, как и условия арифметических задач): составлялись словарные массивы, элементы которых (точнее начальные формы соответствующих слов) отыскивались в «Грамматическом словаре русского языка» А. А. Зализняка [12].

КОД — основная сервисная процедура автоматической генерации морфологических зон словарных статей системы TULIPS-2 для Н-слов и основ И-слов по данным из словаря А.А. Зализняка. В процессе разработки этой процедуры была составлена таблица соответствия словарной информации из словаря А.А. Зализняка словарной информации системы TULIPS-2, которая приводится в приложении 2. Процедура КОД автоматизирует очень трудоёмкую, требующую хорошего знания морфологической модели системы TULIPS-2

работу по составлению словарных статей. Так, при формировании словарной статьи вручную необходимо установить:

- 1) основу (так как в словаре в случае чередований хранится но основа в обычном понимании, а её вариант, встречающийся у большинства форм парадигмы, например: «льв→», а не «лев→»);
- 2) номера М–класса, С–класса и П–класса;
- 3) наличие чередований и их контекст;
- 4) наличие различных исключений и других особенностей словоизменения (см, гл. 2, §2).

Процедура КОД — функция от двух аргументов. Первый аргумент — слово из словаря А.А. Зализняка, второй — заключённые в круглые скобки символы, стоящие около этого слова (правила кодировки этих символов и работы с процедурой КОД см. в приложении 2). Примеры словарных статей: «прихожая (ж(п 4 а))»; «серый (п 1 а)»; «давать (нов 13 в)». По первому элементу словарной статьи из словаря А.А. Зализняка в большинстве случаев определяется номер М–класса, у существительных — также род и одушевлённость, у глаголов — вид. Например, если это «п», то слово относится к 8–му М–классу; «ж» — к 7–му М–классу, женскому роду, неодушевлённое; «нов» — к 9–му М–классу, несовершенному виду. После определения М–класса происходит переход на соответствующую ветвь программы, где по второму элементу — цифре — определяется номер П–класса для каждого конкретного М–класса (напомним, что номера П–классов в словаре А.А. Зализняка и в модели системы TULIPS–2 могут различаться). При несовпадении номеров С–класса и М–класса вторым элементом списка является заключённая в круглые скобки самостоятельная словарная статья (которая и характеризует модель изменения слова). Тогда по первому элементу устанавливается номер С–класса, а из второго определяется М–класс, П–класс и все нужные характеристики слова. Например, из словарной статьи слова «прихожая» — ж(п 4 а) — следует, что это слово относится к 7–му С–классу, но склоняется по модели 8–го М–класса 4–го П–класса. Остальные элементы словарной статьи либо участвуют в определении номера П–класса, либо свидетельствуют о наличии в слове чередований, исключений, об отсутствии каких–либо его форм. Например, помета «И» означает отсутствие у прилагательных кратких форм, «П2» — наличие у слова второго предложного падежа (исключение МИ17 или МИ18), * — наличие чередования. В словаре А.А. Зализняка, как отмечалось выше, любое чередование обозначается *, поэтому при определении его конкретного номера необходимо анализировать структуру слова (последнюю, предпоследнюю букву, концовку основы), а иногда учитывать и некоторые дополнительные пометы в словарной статье, например, букву, обозначающую тип ударения. У глаголов наличие чередования определяется только по цифре из словарной статьи и структуре основы. Например, если на вход процедуры подаются данные: «лев (мо 1 * в)», то по элементу «мо» определяется, что это существительное мужского рода, одушевлённое, по цифре 1 — принадлежность к 1–му П–классу субстантивных слов, по «*» — наличие чередования, по буквам «ле», стоящим перед последней согласной слова определяется номер чередования — 4 (е → ь), буква «в» в данном случае неинформативна. Нужный вариант основы («льв→») определяется по номерам П–класса и чередования. При перекодировании словарной статьи глагола «мыть» — (нсв 12 а) по элементу «нсв» определяется, что это глагол несовершенного вида, по цифре 12 — принадлежность его к 1–му П–классу спрягаемых слов, по последней букве основы — ы — и цифре 12 — наличие чередования 43 (ы → о), по номеру чередования и цифре 12 — нужный вариант основы («мо→»), возвратность определяется по наличию постфиксов «–ся/–сь».

Результатом работы процедуры КОД в большинстве случаев является список, первый элемент которого — основа очередного слова (Н–слово), второй — словарная статья дан-

ной основы (Н–слова) для системы TULIPS–2. Например, для слова «мыть» КОД выдаёт следующий результат:

```
(МО (MS (((1 43 (ТЬ Л ЛА ЛО ЛИ В ВШИ) =МЫТЬ=)))));
```

для слова «кофе»:

```
(КОФЕ (МВ ((7 2 3 0 0) =КОФЕ=))).
```

Для слов, у которых различаются по написанию, например, основы единственного и множественного числа, есть супплетивные формы, неучитываемые чередования, в словаре хранятся два или несколько вариантов основы со специальными ссылками друг на друга (см. гл. 2, § 4). КОД для таких слов выдаёт несколько описанных выше списков (соответственно для каждого варианта основы), заключённых в общие круглые скобки. Например, для слова «цыплёнок» результат такой:

```
(( (ЦЫПЛЕНК (MS (((3 1 (( )) 7110 ((1 1 2 0) (ЦЫПЛЯТ (21 71320)))  
=ЦЫПЛЕНОК=))))))  
(ЦЫПЛЯТ (MS ((21 71320 (7 112 0) ((1 1 1 0) (ЦЫПЛЕНК ((3 1 (( ))  
71110)))) =ЦЫПЛЕНОК=)))))).
```

Для глаголов КОД выдаёт несколько словарных статей, соответственно для форм глагола, форм действительного и страдательного причастий настоящего времени (несовершенного вида) и форм причастия прошедшего времени (для причастий указывается С–класс и значения «дополнительных» ГП — возвратности, вида, времени и залога). Например, для слова «делать» результат такой:

```
(( (ДЕЛА (MS ((1 9120000) =ДЕЛАТЬ=))))  
(ДЕЛАЮЩ (MS ((4 80000 (18 1 2 1 1)) =ДЕЛАЮЩИЙ=))))  
(ДЕЛАЕМ (MS ((1 80000 (18 1 2 1 2)) =ДЕЛАЕМЫЙ=))))  
(ДЕЛАЕМ (MS ((4 80000 (18 1 2 3 1)) =ДЕЛАВШИЙ=))))).
```

Непосредственно записывает в словарь системы результат работы процедуры КОД и производит необходимые обобщения словарных статей для слов с одинаковой основой процедура ЗЗЗ. Работа процедуры проходит в режиме диалога с пользователем. Предложив пользователю (или системщику) вводить словарные статьи из словаря А.А. Зализняка, ЗЗЗ узнаёт, выдавать ли результаты работы процедуры КОД на экран терминала. После окончания ввода словарной информации пользователем процедура сообщает о произведённых ею обобщениях словарных статей, и (по желанию пользователя) печатает список основ, занесённых в словарь системы. Работа процедуры ЗЗЗ выглядит, например, так (сообщения системы печатаются между знаками *):

```
* ПРОЦЕДУРА ПОПОЛНЕНИЯ СЛОВАРЯ * * ВЫДАВАТЬ ЛИ ФОРМИРУЕМЫЕ СЛОВАРНЫЕ СТАТЬИ —  
ДА/НЕТ *  
ДА  
* ВВЕДИТЕ НОВЫЕ СЛОВА. ПРИЗНАК КОНЦА — ( ) *  
  
(ПОРА (Ж 1 А))  
  
* ЗАПИСЫВАТЬ ЛИ В СЛОВАРЬ ПРИЧАСТИЯ? *  
НЕТ  
  
* (ПОР (MS (((1 72200) =ПОРА=)))) *  
(ПОРОТЬ (НСВ 10 С))  
  
* (ПОР (MS (((1 1 (ТЬ Л ЛА ЛО ЛИ В ВШИ) 9120000 МИ15 МИ16) =ПОРОТЬ=)))) *  
(АРМЯНИН (МО 1 « А))
```

```

* ((АРМЯНИН(МС((8 71110((1 1 2 0) (АРМЯН(8 71120)))) =АРМЯНИН=))))
(АРМЯН(МС((8 71120((1 1 1 0) (АРМЯНИН(8 71110)))) =АРМЯНИН=)))) *
( )
* НОВЫЕ СЛОВА ЗАПОМИНАЮТСЯ В ТЕКУЩЕМ СЕАНСЕ *
* НОВЫЕ СЛОВАРНЫЕ СТАТЬИ ДОБАВЛЯЮТСЯ В МАССИВ ГТГ *
* ПРОИЗВЕДЕНО ОБОЩЕНИЕ СЛОВАРНЫХ СТАТЕЙ С ИМЕНЕМ — ПОР *
* ОКОНЧАТЕЛЬНЫЙ ВАРИАНТ ТАКОВ: *
* (МС((1 72200)=ПОРА=((1 1(ТЬ Л ЛА ЛО ЛИ В ВШИ)9120000
МИ15 МИ16)=ПОРОТЬ=))) * * ВЫДАВАТЬ ЛИ СПИСОК ОСНОВ? — ДА/НЕТ *
ДА

* ПОР АРМЯНИН АРМЯН *
* КОНЕЦ СЕАНСА *

```

§ 2. Методы и процедуры автоматической адаптации системы к незнакомым ей языковым объектам.

Средства автоматической адаптации системы на лексическом и морфологическом уровнях позволяют ей самостоятельно установить грамматические характеристики незнакомых языковых объектов, обнаружить и исправить некоторые ошибки в словах (см. §4), выявить специфические для пользователя отклонения от используемой системой модели языка (для слов, записанных в словаре систем) и явно неграмматичные формы слов (см. §4).

При установлении грамматических характеристик незнакомых слов используются: информация, полученная от процедуры MORF, результаты исследования структуры основы и предсказания. Первоначально (в такой постановке задача была предложена 2 года назад М.Г. Мальковским) предполагалось использовать информацию о структуре (морфемном составе) основы с целью установления семантических признаков незнакомых слов. Однако, начатые исследования, во-первых, подтвердили ожидания о принципиальной возможности выявления ряда семантических признаков слов, определяемых их морфемным составом, а, во-вторых, показали, что за решение этой задачи браться еще рано, поскольку без разработки «семантического языка», составления его «лексикона» и правил оперирования с семантическими единицами цели морфемного анализа слишком неопределенны. В то же время оказалось, что, учитывая строение основы, можно существенно повысить точность морфологического анализа незнакомых слов и составляемых для них описаний морфологических признаков [22, 23]. Поэтому в данной работе морфемный состав основы рассматривается лишь с точки зрения выявления морфо-синтаксических признаков, легко фиксируемых в словарной статье (П-класс; значения ГП: вид, род, одушевленность и др.; ожидаемые чередования). Получены также предварительные данные о возможности восстановления некоторых синтаксических свойств слов по их морфологическим признакам и строению основы (например, модели управления).

MORN — основная процедура, реализующая автоматическую адаптацию системы к незнакомым ей словам. На вход процедуры поступает два аргумента: результат работы процедуры MORF для новых и незнакомых слов в виде

$$\left\{ \begin{array}{l} \text{ся} \\ \text{сь} \\ () \end{array} \right\} (< \text{основах флексия}>)$$

$$\begin{aligned} &((\text{М-класс}^1_1 (\text{П-класс} (\text{ы})^1_1 L^1_1) \dots (\text{П-класс} (\text{ы})^1_{n1} L^1_{n1})) \\ &\dots\dots\dots \\ &((\text{М-класс}^k_1 (\text{П-класс} (\text{ы})^k_1 L^k_1) \dots (\text{П-класс} (\text{ы})^k_{nk} L^k_{nk})) \\ &\dots\dots\dots) \end{aligned}$$

и номер слова в предложении. L^i_j — соответствующие данной флексии значения ГП для М-класса i и П-класса (списка П-классов) j , ($i = 1, 2, \dots k; j = 1, 2, \dots n_i$).

В большинстве случаев анализ флексии не позволяет однозначно установить не только П-класс, род существительных, вид глаголов, но даже М-класс словоформ, так как, например, флексия «-Ø» встречается и у глаголов («мог-Ø»), и у существительных («класс-Ø»), и у кратких прилагательных («красив-Ø»), и у слов местоимённого склонения («ядин-Ø»). Для уточнения анализа незнакомых слов MORN использует следующие составляющие основы («диагностический сегмент основы»):

- 1) префикс;
- 2) суффикс или определённое сочетание букв в конце основы;
- 3) последнюю букву основы.

По префиксу можно обнаружить некоторые Н-слова и установить вид некоторых глаголов. Анализ суффикса часто позволяет точно установить М-класс, П-класс, род, а иногда и одушевлённость существительных, вид глаголов, или все нужные характеристики слова. По последней букве основы легко уточняется П-класс, а иногда и С-класс словоформ. В настоящее время MORN учитывает 28 префиксов и 67 суффиксов (при необходимости этот набор можно легко изменить), среди которых 59 — диагностические, т.е. наличие их в основе точно определяет принадлежность слова к какому-то конкретному М-классу (в случае существительных — к конкретному роду) и П-классу; и 8 — полудиагностические, которые могут встречаться у слов, принадлежащих двум различным М-классам, родам существительных или П-классам. Диагностические суффиксы, префиксы, правила уточнения значений ГП основ по их последней букве были получены в результате исследований всех слов, содержащихся в словаре А.А. Зализняка. Почти для каждого суффикса и при учёте префиксов есть слова-исключения, не подчиняющиеся общему правилу.

Для всех диагностических и полудиагностических суффиксов слов-исключений достаточно много (более 400 или около 0,4% всех слов, записанных в словаре А.А. Зализняка), причём учтены только те слова, которые находятся в этом словаре. Записать их все в некоторый специальный словарь системы не всегда целесообразно и не всегда возможно (так как естественный язык — открытая, постоянно расширяющаяся система). В системе TULIPS-2 предусмотрены два режима работы со словами-исключениями, не записанными в словарь системы. В одном режиме при обнаружении в слове диагностического (полудиагностического) суффикса ему приписываются те и только те характеристики, которые определяются этим суффиксом, а в случае отсутствия пересечения значений ГП определяемых флексией и суффиксом в слове ищутся ошибки или начинает анализироваться другой вариант деления слова на основу и флексию (а если это невозможно — слово считается неизменяемым). Действительные характеристики слов-исключений при этом восстанавливаются лишь при диалоге с пользователем. При работе во втором режиме обнаружение диагностического (полудиагностического) суффикса служит лишь для уточнения значений ГП слова, соответствующих значениям ГП, определяемых суффиксом, остальная же словарная информация не уничтожается (как при работе в первом режиме), а передаётся на дальнейшие этапы анализа. Первый режим ориентирован на более точный и быст-

рый анализ основной массы слов, при этом желательно, чтобы слова-исключения были заранее занесены в специальный словарь системы. Второй же режим ориентирован на более правильную работу системы со словами исключениями (например, когда не пересекаются значения Ш, определяемые флексией и суффиксом: «призм-ой» — но флексии «-ой» не может быть у существительных мужского рода, определяемых диагностическим суффиксом «-изм-»).

Отметим также, что все полудиагностические суффиксы на самом деле точно определяют принадлежность слова к конкретному М-классу и П-классу в большинстве его форм или во всех его фордах. Так, например, наборы флексий определяемых суффиксом «-льн-» существительных 12-го П-класса и прилагательных 1-го П-класса не пересекаются.

Учитываемые процедурой MORN составляющие основы, связанная с ними грамматическая информация о незнакомом слове и соответствующие им слова-исключения указаны в приведённых далее таблицах и описаниях.

Используя все указанные возможности для определения грамматических характеристик словоформ, процедура MORN одновременно с анализом незнакомого слова может проводить контроль за правильностью его склонения или спряжения (см. § 4).

Анализ незнакомого слова MORN начинает с варианта с максимальной длиной окончания. Результат её работы сопоставляется с предсказаниями, и, в случае соответствия, запоминается системой. Если же предсказание не подтвердилось, начинает обрабатываться другой вариант деления слова на основу и флексию.

В процедуре MORN предусмотрена возможность повторного обращения к ней с той же основой и той же флексией в случае неподтвердившихся предсказаний или, когда ни один из полученных процедурой MORF вариантов значений ГП не оказался подходящим (см. гл. 2, § 6). Например, словоформа «глаза» может быть воспринята только как существительное мужского рода единственного числа, а есть предсказание, что это существительное множественного числа мужского рода; или словоформа «остриёв» может быть воспринята только как существительное мужского рода, а есть предсказание, что это существительное среднего рода. В подобных случаях делается предположение, что в слове есть исключение, и происходит обращение к процедуре MORN с информацией, связанной с данной флексией, уже, когда она является исключением (если такое возможно). Информацию для повторного обращения к себе процедура MORN готовит сама во время первой обработки данной основы и данной флексии. Если повторная обработка слова закончилась успешно, MORN заносит номера, выявленных исключений в словарную информацию данной основы. Иначе (когда все варианты деления слова на основу и флексию уже просмотрены) в зависимости от предсказаний и режима работы либо слово признаётся неизменяемым, либо в нём ищутся и исправляются ошибки, либо посылается «неуспех» на этап синтаксического анализа, в результате чего может быть изменено предсказание.

Результатом работы процедуры MORN является информация:

а) для этапа синтаксического анализа в виде

(((<номер С-классов> <список значений ГП>) (=R(<основах><М-С-зона>)) <основа>(n i j) . . .),

где «=R» — специальный признак словарной информации для синтеза словоформы, *n* — номер слова в предложении (глобальный параметр), *i* — номер варианта расщепления слова на основу и флексию), *j* — номер трактовки (приписывания различных возможных значений ГП) слова для данного расщепления;

б) для словаря новых слов в виде

((<М-С-зона><основа>) . . .);

- в) оставшиеся необработанными варианты деления слова на основу и флексию с соответствующими возможными значениями ГП;
- г) название поля, определяющего тип объекта (MW, MS или MR);
- д) основа и флексия обрабатывавшейся словоформы.

Например,

ГЕОЛОГ^{MORN} → (MS (ГЕОЛОГ ()) ((7 1 1 1 1) (=R (ГЕОЛОГ (3 71100))) ГЕОЛОГ (1 1 1))) ((3 71100) ГЕОЛОГ) ());

ГОРОДА (с предсказанием, что это существительное мужского рода, множественного числа, именительного падежа)^{MORN} → MS (ГОРОД А) (((7 0 1 2 1) (=R (ГОРОД (1 70100 МИ1))) ГОРОД (1 1 1))) (((1 70100 МИ1) ГОРОД) ());

СОЛИТЬСЯ^{MORN} → (MR (СОЛИ ТЬ) (((9 1 2 (2 6) 0 0 0) (=R (СОЛИ ((1 3) 9120000))) СОЛИ (1 1 1))) (((1 3) 9120000) СОЛИ) (((СОЛИТЬС Я) ((7 ((2 6 7) 0 1 1 (2 4))...))) (0 (СОЛИТЬСЯ ()) ((7 ((1 3 4 5 8 9) 0 1 1 (1 4)) (8 1 1 2 (2 4))...)))))).

Префиксы.

Диагностическими префиксами, учитываемыми процедурой MORN, являются «по-» («по-новому»), «в-» («в-третьих»), «во-» («во-первых»). Наличие в основе этих префиксов является признаком того, что данное слово — Н-слово (наречие или вводное слово). Исключений для этих префиксов нет.

У некоторых глаголов путём анализа префикса можно определить вид. Наличие префикса у таких глаголов означает, что он — совершенного вида, отсутствие — несовершенного вида. Учитываются следующие префиксы глаголов:

вз, вы, до, за, из, ис, на, над, низ, недо, по, об,
от, пере, пре, пред, при, про, под, раз, рас.

Однобуквенные префиксы — в, о, с, у — не учитываются, так как при их рассмотрении в список исключений попали бы все глаголы определённых типов, основа которых начинается на «в», «с», «о», «у».

Любые глаголы с двойными префиксами («понаехать») — совершенного вида. Наиболее часто встречаются следующие двойные префиксы:

навы, напод, повы, пона, вс, пораз, проза.

ТАБЛИЦА ТИПОВ ГЛАГОЛОВ, У КОТОРЫХ МОЖНО УСТАНОВИТЬ ВИД ПО НАЛИЧИЮ ПРЕФИКСА

Концовка инфинитива глагола	Примеры	Исключения
зть	грызть, погрызть	—

Концовка инфинитива глагола	Примеры	Исключения
ить	солить, насолить	долбить, помадить, прекословить, дож- дить, дорожить, безобразить, поганить, полнить, помнить, просить, заботить, бесчестить, портачить, противоречить, прочить, пророчить; √: ·-водить, родить, ·-ходить, ·-возить, женить, ·-носить*
лавать	плавать, сплавать	—
мять	мять, замять	—
ничать	слесарничать, послесар- ничать	—
оть	колоть, наколоть	полоть
ти	нести, отнести	—
чь	беречь, сберечь	—
ыть	плыть, уплыть	—

Суффиксы

ТАБЛИЦА ДИАГНОСТИЧЕСКИХ СУФФИКСОВ

Суффикс	С- класс	П- класс	Род	Оду- шевл.	Вид	Пример	Исключения
ава	9	1	0	—	0	плавать	каравай
ажн	8	1	0	0	—	важный	блажной, важно (как наречие)
анин	7	8	1	1	—	шанин	манганин, цианин, меланин, соланин, софранин, танин, гуанин
ател	7	2	1	0	—	мечтатель	багатель, москатель
атор	7	1	1	0	—	агитатор	—
афи	7	17	2	2	—	география	—
ба	9	1	0	—	0	сгибать	бай, краснобай, пустобай
ащ	8	4	0	0	—	значащий	плащ, пращ
вн	8	1	0	0	—	дивный	древний, ·-давний, фавн, из- давна, сыздавна, королевна, цесаревна, царевна, гривна, поповна, стиховна, ·-давно, равно, неравно, бревно, про- тивно, словно, безусловно, равно

* Здесь и далее признак «√» означает, что данное слово может быть и исключением, и подчи-
няться правилу; уточнение его характеристик нужно производить с использованием, например,
предсказаний. Признак «↔» («↔») означает, что имеются в виду слова, оканчивающиеся на
стоящее после него сочетание букв (и само слово, стоящее после знака «↔»).

Суффикс	С-класс	П-класс	Род	Оду-шев.л.	Вид	Пример	Исключения
граф	7	1	1	0	—	параграф	графы
онок	7	3	1	0	—	цыпленок	—
етел	7	2	1	0	—	свидетель	метель, нетель, петел, добродетель, облетелый
етр	7	1	1	2	—	метр	уретра, метро, гидрометр, осетр
ец	7	5	1	0		комсомолец	кислеца, наглеца, хитреца, дуплецо, письмецо, пальтецо, ружьецо, копьецо, сукнецо, серебрецо, бельецо
етри	7	17	2	2	—	геометрия	заветрие, безветрие, поветрие, эндометрий
еват	8	1	0	0	—	синеватый	—
есят	8	1	0	0	—	десятый	—
ива	9	1	0	—	2	устраивать	закивать, покивать, поспрашивать
ига	9	1	0	—	0	двигать	—
идн	8	1	0	0	—	видный	—кидной, ехидна
изм	7	т	1	2	—	сионизм	клизма, аневризма, призма, кафизма, схизма
ител	7	2	1	0	—	глушитель	канитель, обитель, антитело
истк	7	13	2	1	—	артистка	•—чистка, любисток, свисток, листок
иц	7	15	2	0		приёмщица	лицо, налицо, заподлицо, стойлице, маслице, креслице, пряслице, самолюбьице, здоровьице, ожерельице, —ньице, —тьице
йш	8	4	0	0	—	старейший	гейша, казначейша
ктор	7	1	1	0	—	вектор	—
ла	9	1	0	—	0	лаять	лай (сущ.)
нича	9	1	0	—	0	безобразничать	—

Суффикс	С-класс	П-класс	Род	Одушевл.	Вид	Пример	Исключения
нн	8	1	0	0		данный туманный каменный кухонный	ванна, саванна, канна, манна, панна, осанна, странно, норманн; ранний, сверхранний, искренний, неискренний, выпренный, превыспренный, утренний, предутренний, заутренний, весенний, предвесенний, осенний, предосенний, геенна, антенна, особенно, собственно, следственно, соответственно, естественно, денно, переменно, преимущественно, гребенной, денной, беленной
нн	8	1	0	0	—		зеленной, племенной, межвременной, стремянной, •—семенной, теменной, именной, коренной, подкоренной, куренной, сенной, стенной; —пременно, именно, поименно, несравненно, несомненно, спинной, финна, финн, джинн, белофинн, подлинно, •—сторонний, бонна, донна, мадонна, примадонна, беладонна, колонна, автоколонна, тонна
нт	7	1	1	0	—	аспирант	ванта, веданта, медианта, форманта, доминанта, субдоминанта, варианта, гаранта, куранта, константа, конституанта, инфанта, баранта, бельканто, манто, эсперанто, вента, эвольвента, резольвента, лента, кардолента, кинолента, перфолента, рента, плацента, ленто, пронукиаменто, треченто, квинченто, кватроченто, квинта, пинта, акватинта, анодонта, полфунта, хунта
обн	8	1	0	0	—	способный	надобно, подобно, неудобно, удобно
ова	9	1	0	—	0	действовать	•—кровать
оват	8	1	0	0	—	глуповатый	•—кровать; шлаковата, многовато, жутковато, маловато, тяжеловато, рановато, поздновато, холодновато, трудновато, темновато, страшновато
оги	7	17	2	2	—	биология	—

Суффикс	С-класс	П-класс	Род	Одушевл.	Вид	Пример	Исключения
олог	7	3	1	1	—	идеолог	пологий, гомолог, монолог, полог, аполог, мартиролог, некролог, пролог
онок	7	3	1	0	—	зайчонок	—
ост	7	18	2	2	—	трость	гость, шилохвость, знаменитость, староста, короста, безостый, девяностый, ост, мост, •-хвост, гост, погост, зюйд-ост, периост, намест, помост, диагност, алконост, •-пост, •-рост, тост
разн	8	1	0	0	—	разный	образный, сообразно
сн	8	1	0	0	—	сносный	сосна, кросна, просно, волосной, полосной, переносной, уносной, выносной, набросной, разбросной, сбросной, корпусной, мясной, поясной
сот	8	1	0	0	—	двухсотый	красота, высота, осот, •-сот (числит.)
ств	7	21	3	2	—	свойство	паства, средства, листва, сызмальства, яства, чёрствый
тн	0	1	0	0	—	заметный	цветной, третной, сетной, •-летний, заметно
ува	9	1	0	—	2	раздувать	—
ушк	7	13	2	0	—	девушка	ушко, хлебушко, чадушко
ущ	8	4	0	0	—	будущий	гуща, куща, пуща, пуще
ци	7	17	3	2	—	акация	старушенция, грация, парамеция, технеций, перитеций, экзотеций, лютеций, силиций, ферросилиций, панариции, алириций, патриций, стронций, нунций, сестерций, препуций, кальций
цк	8	3	0	0	—	немецкий	цацка, клёцка
чат	8	1	0	0	—	ступенчатый	девчата, внучата, печать, фотопечать
чив	8	1	0	0	—	изменчивый	сочиво
чик	7	3	1	0	—	лётчик	плечико, личико
чн	8	1	0	0	—	яблочный	мелочный, щелочной, сволочной, ночной, строчной, лучной, случной, ручной, мучной, ключной, кочан

Суффикс	С-класс	П-класс	Род	Одушевл.	Вид	Пример	Исключения
щик	7	3	1	1	—	гонщик	плащик, бензозаправщик, автозаправщик, костылезабивщик, расштыбовщик, бомбардировщик, пикировщик, пищик, –подборщик, снегоуборщик, прыщик, тральщик, ящик, хрящик, ∇: угольщик, подборщик, корректировщик, сортировщик
ыва	9	1	0	—	2	проигрывать	перебивать, побывать
ющ	8	4	0	0	—	знающий	плющ
яни	7	25	3	2	—	расстояние	—
янин	7	1	1	1	—	армянин	—
ятел	7	2	1	1	—	деятель	дятел
ятор	7	1	1	0	—	изолятор	—
ящ	8	4	0	0	—	правлящий	хрящ

ТАБЛИЦА ПОЛУДИАГНОСТИЧЕСКИХ СУФФИКСОВ

Суффикс	С-класс	П-класс	Род	Одушевл.	Вид	Примеры	Исключения
ан	7	25	3	2	—	собрание	германий, •–мания, •–кампания, •–компания, тирания, •–тания, метофания
	9	3	0	—	0	ранить	
енк	7	3	1	0	—	цыплёнок	эвенк
	7	13	2	0	—	плёнка	
ени	7	25	3	2	—	пение	–огения, гардения, нения, лейкопения, шизофрения, –отения, астения, миастения, фоноастения, неврастения, психастения, гений, периселений, алоселений, рений, рутений, парасчений, фений
	9	3	0	—	0	сменить	
ист	7	1	1	1	—	футболист	вист, убиквист, свист, твист, •–лист, автомашинист, аметист, баллиста, триста, циста, дочиста, начисто, монтекристо, монисто, свистеть, свистеть, чистить
	8	1	0	0	—	лучистый	
ик	7	3	1	0	—	чайник	•–дикий, –ликий, калика, елико, клико, трико, колёсико, плечико, личико
	7	13	2	2	—	улитка	

Суффикс	С–класс	П–класс	Род	Оду–шевл.	Вид	Примеры	Исключения
льн	7	12	2	2	—	гладильня	остальной, стальной, шальной, хмельной, •–больной, •–дальний, линкольн, лён, –тельно
	8	1	0	0	—	довольный	
онк	7	3	1	0	—	зайчонок	воронко
	7	13	2	0	—	солонка	
ск	7	3	1	0	—	пропуск	фиаско, каско, альфреско, мяско, –овской, складской, слободской, •–заводской, •–городской, чудской, людской, рядской, мужской, чухломской, сумской, ямской, двинской, донской, поварской, писарской, мастерской, кучерской, мирской, •–морской, хуторской шегольской, день–деньской
	7	13	2	0	—	миска	
	8	3	0	0	—	русский	

ТАБЛИЦА УЧИТЫВАЕМЫХ СУФФИКСОВ ДЛЯ РАЗЛИЧНЫХ М–КЛАССОВ И РОДОВ СУЩЕСТВИТЕЛЬНЫХ

М–Класс 7			М–класс 8	М–класс 9
Мужской род	Женский род	Средний род		
Диагностические суффиксы				
анин	афи	ств	ажн	ава
ател	етри	яни	ащ	ба
атор	истк		вн	ива
граф	иц		еват	ига
енок	оги		есят	ла
етел	ост		идн	нича
етр	ушк		йш	ова
ец	ци		нн	ува
изм			обн	ыва
ител			оват	
ктор			разн	
нт			сн	
олог			сот	
онок			тн	
чик			ущ	
щик			цк	
янин			чат	
ятел			чив	
ятор			чн	
			ющ	
			ящ	

М–Класс 7			М–класс 8	М–класс 9
Мужской род	Женский род	Средний род		
Полудиагностические суффиксы				
енк	енк	али	ист	ани
ик	ик	ени	льн	ени
ист	льн		ск	
онк	онк			
ск	ск			

Уточнение значений ГП словоформ по последней букве основы (ПБО).

I. Если у словоформы пустая флексия и ПБО гласная, ь или й, то это неизменяемое слово («кофе», «однако», «какаду»).

II. Существительные мужского рода.

1. Если ПБО а, е, о, й, у или к, то П–класс может быть только 6.
2. Если ПБО и, то П–класс может быть только 7.
3. Если ПБО ц, то П–класс может быть только 1 или 5.
4. Если ПБО ш, щ, ч или д, то П–класс может быть только 4 или 9.
5. Если ПБО к, г или х, то П–класс может быть только 3.
6. Основы, оканчивающиеся на не указанные выше буквы, могут относиться только к П–классам 1, 2 или 8.

III. Существительные женского рода.

1. Если ПБО а, е, о, й, у или ы, то П–класс может быть только 16.
2. Если ПБО и, то П–класс может быть только 17.
3. Если ПБО ц, то П–класс может быть только 11 или 15.
4. Если ПБО ш, щ, ч или ж, то П–класс может быть только 14 или 19.
5. Если ПБО к, г или х, то П–класс может быть только 13.
6. Основы, оканчивающиеся на не указанные выше буквы могут относиться только к П–классам 11, 12 или 18.

IV. Существительные среднего рода.

1. Если ПБО и, то П–класс может быть только 25.
2. Если ПБО ц, то П–класс может быть только 21 или 23.
3. Если ПБО щ, то П–класс может быть только 23.
4. Если ПБО к, то П–класс может быть только 21.
5. Если ПБО м или н, то П–класс может быть только 21 или 27.
6. Если ПБО л, р или в, то П–класс может быть только 21.
7. Если ПБО ь, то П–класс может быть только 28.
8. На другие буквы алфавита основы существительных среднего рода оканчиваться не могут.

V. Слова адъективного склонения.

1. Если словоформа имеет флексию «–е» (т.е. С–класс 17), то её основа может оканчиваться только на ч, ж или ш; при этом наличие префикса «по–» — признак слабой сравнительной степени прилагательных («поярче»), отсутствие — сильной сравнительной степени («ярче»). Исключение составляют прилагательные с основой на «по–» («пологий»).
2. ПБО не может быть гласной, ь, и, ц, за исключением слов типа «длинношей», которые относятся к 6–му П–классу.
3. Если ИБО к, г или х, то П–класс может быть только 3 или 8.

4. Если ПБО ж, щ, ш или ч, то П–класс может быть только 4 или 8.
5. Если ПБО ц, то П–класс может быть только 5.
6. Основы, оканчивающиеся на не указанные выше буквы, могут относиться только к П–классам 1, 2, 7.
7. Перед постфиксом «–ся» может стоять только м, х, й или гласная, при этом ПБО может быть только ш или щ (П–класс 4).

VI. Слова местоименного склонения.

1. Если ПБО н или в, то П–класс может быть только 1.
2. Если ПБО и или ь, то П–класс может быть только 2, причём в основе есть чередование «и–ь» (№6) при флексии «–и».
3. На другие буквы алфавита основы слов местоименного склонения оканчиваться не могут.

VII. Глаголы.

1. Перед постфиксом «–ся» может стоять только согласная.
2. Перед постфиксом «–сь» может стоять только а, е, и, о, у, е или я, причём при флексии «–а» (флексия деепричастия) ПБО может быть только щ, ж, ч или щ.
3. При флексиях «–еть» и «–ев» (П–класс 9) ПБО может быть только р.
4. При флексии «–ь» (П–класс 8) ПБО может быть только ч.
5. При флексии «–та» (П–класс 7) ПБО может быть только 3 или с.
6. При флексиях «–л», «–в» и «–вши» ПБО могут быть только а, я, у, е, о, ы или и. При этом:
если ПБО а, то П–класс может быть только 1, 2, 4 или 6;
если ПБО я, то П–класс может быть только 1, 3 или 6;
если ПБО у, то П–класс может быть только 1 или 2;
если ПБО е, то П–класс может быть только 1, 3 или 5;
если ПБО о, то П–класс может быть только 1;
если ПБО ы, то П–класс может быть только 1 или 6;
если ПБО и, то П–класс может быть только 1, 3, 4 или 5.
7. При флексии «–ть» ПБО могут быть все гласные, указанные в пункте 6 (с соответствующими ограничениями), а также з и с (П–класс 7).
8. При флексии «–Ø» ПБО могут быть только г, к, р, з, с, б, п или х. При этом:
если ПБО г или к, то П–класс может быть только 2 или 8;
если ПБО р, то П–класс может быть только 9;
если ПБО з, с или б, то П–класс может быть только 2, 6 или 7;
если ПБО л или х, то П–класс может быть только 2.
9. При флексиях «–ла», «–ло» и «–ли» ПБО могут быть все буквы, указанные в пунктах 6 и 8 (с соответствующими ограничениями).
10. При флексии «–ши» ПБО могут быть только з, с, б, г, к или р. При этом:
если ПБО з, с или б, то П–класс может быть только 7;
если ПБО г или к, то П–класс может быть только 8;
если ПБО р, то П–класс может быть только 9.
11. При остальных флексиях глаголов («–у», «–ю», «–ишь», «–ешь», «–ит», «–ет», «–им», «–ем», «–ите», «–ете», «–ут», «–ют», «–ат», «–ят», «–а», «–я» «–и», «–й», «–йте»):
если ПБО ь, я, и, ю или у, то П–класс может быть только 1;
если ПБО а, е, л или о, то П–класс может быть только 1 или 3;
если ПБО я, ш, щ или ч, то П–класс может быть только 2, 4 или 5;
если ПБО з, с или д, то П–класс может быть только 5, 6 или 7;
если ПБО г или к, то П–класс может быть только 8;
если ПБО в или м, то П–класс может быть только 3 или 6;
если ПБО б или т, то П–класс может быть только 3, 5 или 7;

- если ПБО п, то П–класс может быть только 3;
- если ПБО ф, то П–класс может быть только 3;
- если ПБО н, то П–класс может быть только 2, 3 или 6;
- если ПБО р, то П–класс может быть только 3, 5, 6 или 9. 12.

На другие буквы алфавита при каждой конкретной флексии основы глаголов оканчиваться не могут.

Дополнительные правила, используемые при анализе незнакомых слов.

1. Существительные среднего рода — неодушевлённые. Исключения: чудовище, чудище, козлище, идолище, дружище, уродище, великанище, парнище, котище, комарище, плутище, дурачище, человечье, мужичье, волчище, бычище; V: создание, приведение, детище, ничтожество, божество, существо, лицо.
2. Предпоследняя буква основы существительных среднего рода не может быть а, я, у, ю, э, ы, и; исключения составляют слова, основа которых оканчивается на л, в, иц или ищ.
3. Формы настоящего времени могут быть только у глаголов несовершенного вида, будущего — только у глаголов совершенного вида.
4. Глаголы на «-нуть» и «-ать» — совершенного вида. Исключения: класть, красть, слабнуть, гибнуть, зябнуть, дрябнуть, гнуть, блекнуть, никнуть, жолкнуть, молкнуть, мокнуть, меркнуть, горкнуть, тускнеть, мякнуть, сякнуть, тонуть, слепнуть, крепнуть, липнуть, сипнуть, терпнуть, гаснуть, виснуть, киснуть, пахнуть, чахнуть, тихнуть, дохнуть, глохнуть, сохнуть, тухнуть, дряхнуть, стыннуть, льнуть; V: бухнуть.
5. У существительных мужского рода единственного числа и у всех существительных во множественном числе совпадают следующие формы: у одушевлённых существительных — формы родительного и винительного падежей, у неодушевлённых — формы именительного и винительного падежей.

§ 3. Обучение системы пользователем.

Обучение системы пользователем осуществляется с помощью процедуры ОБП. Обучение проходит в режиме диалога. После вызова соответствующих процедур система предлагает пользователю сообщить информацию, которая должна быть занесена в долговременную память; и пользователь начинает вводить свои обозначения, сокращения, новые термины, указывая их нужные характеристики. Отметим, что никаких дополнительных знаний языка, кроме обычных школьных, при этом не требуется. Если система не поняла какое-то значение грамматической характеристики, введённое пользователем с терминала, или он указал не все нужные данные, то она переспрашивает его о нужных значениях. Когда пользователь заканчивает ввод информации, ОБП формирует словарные статьи новых языковых объектов, производя необходимые обобщения (это бывает крайне редко и только в случае терминов) и записывает их в зону данного пользователя. По желанию пользователя список вновь введённых языковых объектов может выдаваться на экран терминала,

Если, например, пользователь для удобства хочет обозначить словосочетание «математический анализ» как «МА», он, вызвав нужные процедуры, задаёт следующие данные:

МА = МАТЕМАТИЧЕСКИЙ АНАЛИЗ (СУЩ МУЖ НЕОД) .

При этом в зоне данного пользователя будет записана такая словарная статья для «МА»:

МА → (MW(((7 210 0)=МАТЕМАТИЧЕСКИЙ АНАЛИЗ=))).

Если же пользователь хочет ввести новый термин, например, «флексия», он может задать информацию в таком виде:

ФЛЕКСИЯ = (СУЩ ЖЕН НЕОД) .

Информацию пользователь может задавать и на обычном русском языке, например, так:

СЛОВО ФЛЕКСИЯ — НЕОДУШЕВЛЁННОЕ СУЩЕСТВИТЕЛЬНОЕ
ЖЕНСКОГО РОДА.

Форма записи грамматической информации о слове выбирается по усмотрению пользователя.

ОБП преобразует полученные данные в предсказание, в данном случае — в (7 2 2 11), т.е. анализируемое слово принадлежит 7-му М-классу, женского рода, неодушевлённое, единственного числа, именительного падежа. Затем с этим предсказанием происходит обращение к процедуре WANL (см. гл. 2, §6). При этом режим работы WANL — «выдать только предсказанный результат». После возможного уточнения номера П-класса результат работы процедуры WANL записывается в словарь системы в зону данного пользователя (или проблемной области).

Для слов с особенностями склонения (спряжения) указываются дополнительные характеристики в конце списка значений ГП. Например,

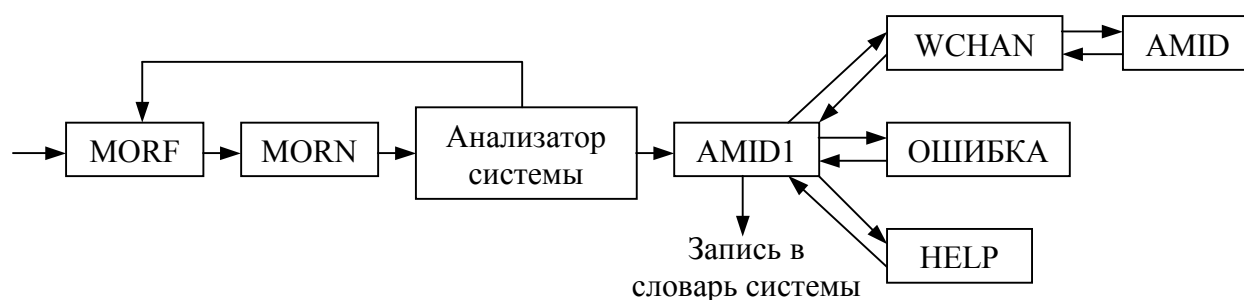
САНИ = (СУЩ ЖЕН НЕОД (Ш)) ;
СКАЗУЕМОЕ = (СУЩ СР НЕОД (ПРИЛ)) и т.п.

Большие массивы словарной информации пользователь также может вводить с помощью процедуры 333 (см. §1).

Отметим, что обработка и запоминание ответов пользователя на запросы системы является обучением системы пользователем только в том смысле, что он — источник информации о языковом объекте, но обучение здесь происходит по инициативе системы.

§ 4. Самообучение системы.

Процесс автоматического самообучения системы схематически можно представить следующим образом:



В течение сеанса работы в анализаторе системы собираются СПИСОК незнакомых основ И-слов и Н-слов — S — и список соответствующих им значений ГП (словарная информация, полученная процедурой MORN) — V. После окончания очередного сеанса работы системы эти списки передаются процедуре AMID1 для дальнейшей обработки, обобщения и запоминания. AMID1 управляет работой процедур WCHAN, AMID) и HELP.

Процедура WCHAN — функция от двух аргументов: первый — список S в виде

$$((\langle \text{основа}_1 \rangle \langle \text{флексия}_1 \rangle) \dots (\langle \text{основа}_n \rangle \langle \text{флексия}_n \rangle));$$

второй — список V в виде

$$(((\langle \text{М-С зона}_1^1 \text{ основа}_1^1 \rangle)_1 \dots) \dots ((\langle \text{М-С зона}_n^1 \text{ основа}_n^1 \rangle) \dots)_n).$$

Сначала WCHAN находит в S одинаковые основы, и с помощью процедуры AMID производит, если это возможно, обобщение значений ГП в соответствующих им элементах из списка V, а также вычёркивает из S одну из основ. Работа процедуры AMID заключается в том, что она находит пересечение значений ГП элементов списка V, заносит в полученную словарную статью основы нужные номера чередований и исключений. Если в словарной информации одной из признанных тождественными основ занесён какой-либо номер чередования, а в другой — нет, то при обобщении двух элементов из V в получившейся словарной статье чередование остаётся. Если в словарной статье одной из одинаковых основ записано некоторое чередование, а в словарной статье второй — другое чередование, то при их обобщении в получившуюся словарную статью записываются оба чередования (ошибки при этом быть не может, так как существует проверка на совпадение номеров С-класса, П-класса, на соответствие флексий). Аналогично процедура работает и в случае наличия исключений. Отметим, что обобщение производится лишь в случае пересечения значений ГП двух одинаковых основ. Если же рассматриваемые основы относятся к двум разным словам (например, «поле» и «пел»), то объединение их словарных статей происходит уже при записи в словарь системы. Обобщая словарные статьи, соответствующие одинаковым основам, процедура AMID может и ошибиться. Так, слова «строй» и «страх» могут быть восприняты как формы слова с основой «стр-» (по аналогии с «гор-ой» и «гор-ах»). Поскольку таких слов немного, целесообразно их заранее занести в специальный словарь системы.

Обработав все слова с одинаковой основой, процедура WCHAN проверяет, есть ли в списке S основы, которые могут соответствовать сильной и слабой сравнительной степени одного прилагательного (различающиеся только наличием префикса «по-» — «краснее» и «покраснее»); и также вычёркивает из списка S основу с префиксом «по-» и с помощью AMID обобщает соответствующие значения ГП из списка V. отождествить две основы «незаконно» практически невозможно, так как происходит проверка на совпадение номеров С-класса, П-класса, на флексию (в данном случае возможны только флексии «-е», «-ее», «-ей»), и если есть какого-либо несоответствие, обобщение соответствующих данных не происходит.

Более сложно обнаружить тождественность основ, связанных чередованием. Чтобы проверить возможность чередования букв в двух основах, надо найти те части основ, которыми они различаются, и проверить, не могут ли они быть «связаны» одним из учитываемых системой чередований? Если да, то при допустимой ли флексии оно применяется и при допустимом ли контексте (в каком месте основ находятся несовпадающие буквы, какие у них необходимые «соседи» слева и (или) справа)? Возможно ли предполагаемое чередование при каком-нибудь из определённых процедурой MORN М-классов и П-классов?

Если же какие-то условия применимости чередования не выполнены, обобщение значений ГП этих основ не происходит. Но, когда WCHAN установит наличие чередования, она может значительно уточнить значения ГП основы с чередованием, так как каждое конкретное чередование может быть в основах лишь определённого М-класса и определённого П-класса, и в большинстве случаев неясными остаются только вид глаголов или одушевлённость существительных. При обобщении значений ГП основ, связанных чередованием, в словарную информацию автоматически заносится нужный вариант основы, номер чередования и его контекст, а из списка S вычёркивается определённый вариант основы с чередованием.

Результатом работы процедуры WCHAN являются изменённые и укороченные (в результате описанных выше обобщений) списки S и V. Например,

```

S = ((ТУРИСТ А) (ТУРИСТ У) (ИЗМЕНЧИВ ЕЕ) (ПОИЗМЕНЧИВ ЕЕ)
      (ЦЫПЛЁНКОКК ( )) (ЦЫПЛЁНК ОМ) (ЖАР ОЙ) (ЖАР ИТ))
V = (((1 71100)=W=) (((1 80000)=W=)
      (((1 71100)=W=) (((1 80000)=W=)
      (((1 80000)=W=) (((3 70100)=W=)
      (((3 70100)=W=) (((11 70200)=W=)
      (((1 7) 80000)=W=) (((3 5)9100000)=W=)))
      WCHAN→
S' = ((ТУРИСТ А) (ИЗМЕНЧИВ ЕЕ) (ЦЫПЛЁНК ОМ) (ЖАР ОЙ) (ЖАР ИТ))
V' = (((1 71100)=W=) (((1 80000)=W=) (((3 1( )) 70100)=W=)
      (((11 70200)=W=) (((1 7)8000)=W=)
      (((3 5)9100000)=W=)))

```

По спискам, полученным от процедуры WCHAN, процедура AMID1 формирует словарные статьи новых слов, производя необходимые обобщения (аналогично процедуре 333) и записывает их в словарь системы. Зона, в которую записывается новое слово, определяется по результатам обращения к метауровню. Чтобы устранить оставшиеся неоднозначности в словарной информации (например, точно не определён П-класс, вид глагола) AMID1 обращается к процедуре HELP, которая просит разъяснений у пользователя. При разработке процедуры были проведены исследования и составлены описания возможных неоднозначных ситуаций при каждой конкретной флексии, которые не может устранить процедура MORN, анализируя структуру основы, и списки соответствующих этим ситуациям вопросов пользователю. Вопросы подбирались таким образом, чтобы ответы на них несли максимум нужной информации. Важно, что для ответа системе пользователь не должен знать ни её морфологическую модель, ни структуру словаря. Запросы системы выглядят, например, так: «Какого рода существительное <A>? — М/Ж/С», «Существительное <A> — одушевлённое? — ДА/НЕТ», «Слово — глагол? — ДА/НЕТ». «Какого вида, совершенного (что сделать? — 1) или несовершенного (что делать? — 2) глагол <C>? — 1/2», «Приведите начальную форму слова <W>» и некоторые другие.

При одних флексиях (»-ее», «-ий») все нужные значения ГП однозначно определяются ещё при работе процедуры MORN, при других — неопределёнными остаются либо только вид глаголов (»-ши»), либо только одушевлённость существительных (»-ью», «-ев»), либо род, П-класс и одушевлённость существительных (»-ами», «-ям»), либо М-класс и некоторые значения ГП для каждого М-класса.

На вход процедуры HELP поступают: основа, флексия анализируемой словоформы и предварительная словарная информация для данной основы, где вместо конкретных значений ГП может стоять «∅», и может быть несколько интерпретаций основы, HELP определяет наличие неоднозначностей в полученной информации и по флексии выбирает вопросы, которые необходимо задать пользователю. Проанализировав ответ пользователя и, изменив соответствующим образом словарную статью, HELP опять проверяет, остались

ли ещё неуточнённые значения ГП слова. Если да, то опять задаёт пользователю нужный вопрос, если нет, то заканчивает свою работу. Результатом работы процедуры HELP является список, состоящий из номера М-класса обрабатываемого слова и морфологической зоны его словарной статьи.

§ 5. Автоматическое исправление ошибок во входном тексте на естественном языке.

При работе за терминалом (как и в обычной речевой практике) люди часто допускают случайные или специфические для их речи ошибки. Способность системы, обеспечивающей диалог человека с ЭВМ на естественном языке, автоматически обнаруживать и исправлять ошибки во входном тексте является очень важной, так как экономит время пользователя, время центрального процессора, повышает устойчивость работы системы в целом, делает диалог с машиной удобным и естественным [22],[24]. Поняв ошибочный вариант какого-либо слова, синтаксической конструкции, система, имеющая средства автоматического исправления ошибок, выполняет нужные исправления, не прекращая анализ текста (т.е. «пропускает» ошибочный вариант на последующие этапы обработки). Помимо ошибок, характерных для текстов на искусственных языках (пропуск, замена букв и т.п.), во входном тексте на естественном языке могут встретиться и различные неграмматические («лектора») или вариантные («туфель» — «туфля») формы и конструкции. Проблема автоматического исправления ошибок существенно усложняется также в связи с тем, что в системе попользуется открытый словарь. В некоторых случаях без помощи пользователя практически невозможно отличить правильную форму незнакомого слова от формы слова, известного системе, но встретившегося в тексте с ошибкой. При этом если пытаться воспринять любое незнакомое слово как форму известного слова с ошибкой, то система будет эффективно работать в случае случайных ошибок. Если же считать, что ошибок быть не должно, тогда проще обрабатывать незнакомые слова. Какой из указанных режимов работы выбрать и в каких ситуациях, как их чередовать и совмещать установит практика. Практика также определит оптимальный уровень самостоятельности системы при обработке незнакомых слов. При практическом использовании системы в одних проблемных областях (например, при общении студентов с автоматизированной обучающей системой) не страшно, если система неправильно воспримет какое-либо слово; а в других (например, при решении задач оперативного управления) ошибки допускаться не должны, и в сомнительных ситуациях система обязана разрешать неоднозначность только с помощью пользователя.

На лексическом и морфологическом уровнях ошибки можно обнаружить как в словах, известных системе, так и в словах, встретившихся ей впервые. Контроль за правильностью изменения слов, занесённых в словарь системы, проводит процедура STFL (см. гл. 2, § 6, стр. 31). Она обнаруживает ошибки, связанные с невыполнением или неправильным применением правил чередования («исчезнул» вместо «исчез») и исключений («якори» вместо «якоря»), используя при этом предсказания и данные из словарной статьи основы. STFL также учитывает допустимые в русском языке колебания в роде («рельс» — «рельса») и одушевлённости («выделить вирус» — «выделить вируса») существительных [25]. В подобных ситуациях процедура временно изменяет значения соответствующих показателей в словарной статье данной основы. Ошибочные и вариантные формы слов запоминаются в зоне данного пользователя, так как предполагается, что те же формы могут встретиться в его речи и в дальнейшем (таким образом, учитываются речевые особенности пользователя). Например, если системе встретился ошибочный вариант известного ей слова «лектор» — «лектора» (число и падеж можно определить, учитывая контекст), тогда к словарной информации основы «лектор» будет добавлено исключение МИ1 и изме-

нённая словарная статья будет записана в зону данного пользователя. Но если бы слово «лектор» было незнакомо системе, то соответствующая словарная статья с исключением МИ1 была бы записана в общую зону.

Ошибки при склонении и спряжении незнакомых слов обнаруживает и исправляет процедура MORN (см. § 2). Исследуя структуру основы, MORF во многих случаях может точно установить П–класс словоформ, а, значит, и проверить, возможна ли данная флексия у слов данного П–класса. Если такой флексии у слов данного П–класса быть не может, то MORN исправляет её на нужную флексию (например, «–ит» на «–ет», «–ому» на «–ему»). Прежде, чем произвести исправления, MORN спрашивает у пользователя, согласен ли он (так как возможно неправильное расщепление основы на основу и флексию или ошибка в основе). Когда пользователь согласен, MORN вместе с исправлением окончания меняет и связанную с ним грамматическую информацию, например, тип спряжения при замене «–ишь» на «–ешь», тип склонения при замене «–ая» на «–яя». Например, если процедуре встретилось слово «наступишь», то по концу основы «–па–», она определяет первое спряжение глагола и исправляет флексию «–ишь» на «–ешь». Если встретилось слово «глубокий», то по последней букве основы «к» MORF определяет принадлежность прилагательного к 3–му П–классу и заменяет флексию «–ый» на «–ий». Отметим, что в процедурах STFL и MORN исправление ошибок (как и учёт колебаний) при словоизменении происходит параллельно с морфологическим анализом слова и требует очень мало машинного времени.

Труднее найти ошибку в основах новых слов или слов, известных системе, но неверно набранных пользователем на клавиатуре терминала. Основные классы учитываемых случайных ошибок таковы:

1. пропуск одной буквы («масив» вместо «массив»);
2. одна лишняя буква («жаанал» вместо «канал»);
3. замена одной буквы («ПЛЕНЕР» вместо «ПЛЭНЕР»);
4. перестановка двух рядом стоящих букв («сторка» вместо «строка»);
5. «дисплейные» ошибки, возникающие при несвоевременном переключении регистров клавиатуры терминала («ЛИСП» вместо «ЛИСП»), сбое на линии связи с ЭВМ.

В системе также считается, что первая буква основы безошибочна и в слове может быть допущена только одна ошибка. Исключение составляют дисплейные ошибки: они обнаруживаются и в первой букве основы, и в нескольких её позициях («START» вместо «СТАРТ»), Указанные ограничения и классификация случайных ошибок приняты с учётом результатов исследований речевых ошибок из [26], исследований проблемы нейтрализации ошибок во входных текстах на формальных языках [27], наблюдений за ошибками лиц, работающих за терминалом (были проведены исследования случайных ошибок, допущенных при наборе текстов некоторых инструкций и описаний за терминалом) и факта наличия в русском языке большого количества слов, различающихся только двумя буквами («дека» — «доска», «класс» — «квас», «белый» — «бедный», «полоскать» — «потаскать») или только первой буквой основы («дом» — «ком» — «ром», «ночка» — «дочка» — «кочка» — «бочка» — «почка» — «точка», «пить» — «лечь» — «вить» — «жить» — «бить» — «шить»).

Обнаруживает случайные ошибки в основах слов процедура ОШИБКА. Обращение к ней происходит из процедур AMID1 или WANL когда есть предположение (например, в предсказаниях указана конкретная основа), что данное слово — искажённая форма уже встречавшегося системе слова. ОШИБКА — функция от одного аргумента: списка различных

основ. Для отождествления основ при наличии ошибок процедура каждой из них ставит в соответствие целочисленный вектор из 32 элементов

$$A = (a_1, \dots, a_{32}),$$

где a_i — число вхождений в данную основу i -той буквы алфавита (исключая «ё»), $i = 1, \dots, 32$.

Для компактного хранения таких векторов специальные процедуры взаимнооднозначно представляют их в виде трёх одиннадцатиразрядных чисел (т.е. для каждой основы такой вектор занимает три машинные ячейки).

На множестве векторов введена метрика (в обычном математическом понимании):

$$M(A_1, A_2) = \sum_{i=1}^{32} |a_i^1 - a_i^2|.$$

Две основы считаются тождественными, если расстояние между соответствующими им векторами не больше двух, причём равенство возможно только в том случае, когда длины двух основ совпадают. Предлагаемая метрика — лишь одна из возможных. Она строилась с учётом ограничений на типы встречающихся ошибок (т.е. все типы случайных ошибок, учитываемых в системе, с помощью этой метрики обнаруживаются, а слова с любыми другими отклонениями отождествляться системой не будут) и, как показали эксперименты, работает достаточно эффективно. Результатом работы процедуры ОШИБКА являются пары основ, в которых одна основа может быть получена из другой в результате какой-нибудь из пяти типов ошибок.

Примеры работы процедуры ОШИБКА:

длинный длинный длинный) ^{ОШИБКА} → ((длинный длинный)
 (длинный длинный)
 (длинный длинный))
 (ЛАПКА ПАПКА ЛПКА) ^{ОШИБКА} → ((ЛАПКА ЛПКА))

Отметим, что ошибки ищутся только среди тех основ, у которых есть общие значения ГП. Процедура ОШИБКА может и сама ошибиться. Например, основы слов «страница» и «странница» могут быть отождествлены.

Заключение

Предложенная работа завершает важный этап решения сложной и актуальной проблемы создания программных средств, обеспечивающих диалог человека с ЭВМ на естественном языке — этап разработки (на лексическом и морфологическом уровнях) методов, принципов и алгоритмов адаптации и обучения на языковом уровне. Использование разработанных средств адаптации к незнакомым языковым объектам и конструкциям, обучения языку, автоматического исправления случайных ошибок во входном тексте в системах обеспечения диалога может повысить «устойчивость» и надёжность работы таких систем, сделать диалог для пользователя естественным и необременительным. В частности получены следующие результаты:

1. Разработаны методы программной реализации систем, обеспечивающих адаптивный диалог человека с ЭВМ на естественном языке, и представления знаний о естественном языке*.
2. На основе предложенных методов и принципов разработана и программно реализована полная модель русского словоизменения, ориентированная на анализ и синтез словоформ над «открытым» словарём основ и неизменяемых слов, реализованы программные средства сопровождения словаря. С практической точки зрения важно, что разработанная модель «согласована» с известным фундаментальным «Грамматическим словарём русского языка» А.А. Зализняка, а её программный компонент включает процедуры автоматического отображения словарных статей из этого словаря в словарные статьи модели. Для лексического и морфологического уровней разработаны методы и алгоритмы: автоматизированного пополнения базы знаний о русском языке; автоматической адаптации к незнакомым системе языковым объектам и индивидуальным особенностям пользователя в употреблении языка (в том числе восстановления морфологических признаков незнакомых слов по результатам исследования строения основы, грамматическим признакам, связанным с окончанием, и контексту); обучения системы пользователем и самообучения; автоматического исправления ошибок во входном тексте на естественном языке.

Из результатов работы следует, что предложенная модель является реальной практической моделью, созданной на основе фундаментальных лингвистических исследований, результаты работы которой проверены на реальном и обширном фактическом материале.

Проведённые исследования показали также, что можно восстанавливать и некоторые синтаксические и семантические признаки слов, извлекая информацию только из текста. Уже предприняты попытки [28] автоматического восстановления синтаксических свойств слов по их морфологическим признакам, структуре основы и контексту. Продолжение этих исследований, включающее разработку аппарата адаптации и обучения системы на синтаксическом уровне, рассматривается как одно из основных направлений работы в ближайшие годы.

Предстоит также большая и сложная работа по разработке и развитию метауровня языковых знаний системы. В нынешнем варианте знания метауровня представлены только в виде процедур, что затрудняет их изменение (как автоматическое, так и пользователем), а обучение на метауровне — это ещё более высокая ступень обучения, которую в перспективе предстоит разработать.

Как уже отмечалось в основном тексте, хотя предлагаемые в работе алгоритмы и процедуры реализованы для экспериментальной системы искусственного интеллекта TULIPS–2, начаты работы и по их практическому использованию [29]. Планы хозяйственных исследований на 1982 г. и последующие годы предполагают активное использование полученных результатов в задачах прикладной лингвистики и при создании практических систем обеспечения диалога человека с ЭВМ на естественном языке, в том числе и вне работ над системой TULIPS–2. Описания морфологической модели, вынесенные в приложения, независимы от системы TULIPS–2, и представлены, в частности для того, чтобы ими можно было воспользоваться.

Автор глубоко признателен за общее руководство, ценные советы и замечания своим научным руководителям Михаилу Георгиевичу Мальковскому и Николаю Павловичу Три-

* Результат получен совместно с М.Г. Мальковским.

фонову. Автор также благодарит всех, принимавших то или иное участие в обсуждении текста диссертации и затрагиваемых в работе вопросов — прежде всего В.Н. Пильщикова; студентов ВМиК МГУ Н.В. Винокурову и Т.В. Грацианову, помогавших в отладке и тестировании отдельных программ и перепрограммировании ряда алгоритмов на языке ПЛ/1 (ОС ЕС) в рамках хоздоговорных исследований.

Литература

- [1] **Кулагина О.С.** Исследования по машинному переводу. — М.: Наука, 1979.
- [2] Синтаксический компонент в системах машинного перевода. Обзорная информация. ВЦП, Сер. 2, Машинный перевод и автоматизация информационных процессов. Изд-во ВИНТИ, М., 1981.
- [3] **Мальковский М.Г., Волкова И.А.** Анализатор системы TULIPS-2. Морфологический уровень. — Вестн. Моск. ун-та. Сер. Вычисл. матем. и киберн., 1981, № 1, стр. 70-76.
- [4] **Виноград Т.** Программа, понимающая естественный язык. — М.: Мир, 1976.
- [5] **Мальковский М.Г.** Программа APRIL, решающая арифметические задачи в словесной формулировке. — В кн.: Алгоритмы и алгоритмические языки. Вып. 6. — М.: Изд. ВЦ АН СССР, 1973, стр. 113-159.
- [6] **Шенк Р.** Обработка концептуальной информации. — М.: Энергия, 1980.
- [7] **Попов Э.В.** Система взаимодействия с ЭВМ на ограниченном русском языке. Программирование, 1978, №4, стр. 66-76.
- [8] **Абрамов В.Г., Брябрин В.М.** и др. ДИЛОС — диалоговая система для взаимодействия с ЭВМ на естественном языке. — М.: Изд. ВЦ АН СССР, 1979.
- [9] **Микулич Л.И., Червоненкис А.Я.** Специализированная диалоговая система. — В кн.: Вопросы разработки прикладных систем, Новосибирск: ВЦ СО АН СССР, 1979, стр. 111-129.
- [10] **Мальковский М.Г.** Программа, понимающая естественный язык. Автореф. канд. дисс. — М.: МГУ, 1981.
- [11] **Гладун В.П., Зверев Ю.М., Яворский А.Л.** Обучаемая система анализа и синтеза текстов естественного языка. — В кн.: Диалог в автоматизированных системах. Материалы семинара. — М.: Изд. МДНТП, 1981, стр. 45-51.
- [12] **Зализняк А.А.** Грамматический словарь русского языка. — М.: Русский язык, 1977.
- [13] **Белоногов Г.Г., Губарь Н.Т., Новосёлов А.П.** Морфологический анализ слов на основе словаря словоформ. — М.: НТИ, ВИНТИ, Сер. 2. 1975, № 9. стр. 29-37.
- [14] **Бидер И.Г., Большаков И.А., Еськова Н.А.** Формальная модель русской морфологии, Часть 1-2, ИРЯ, Предварительные публикации, Вып. III, 112, — М., 1978.
- [15] **Гончаров О.А.** АДС — средство производства диалоговых программ. — В кн.: Диалог в автоматизированных системах. Материалы семинара. — М.: Изд. МДНТП, 1981, стр. 16-23,
- [16] **Сенин Г.В.** От анализа естественного языка к адаптивному диалогу. — В кн.: Семинар проекта «ДИАЛОГ». Тезисы докладов, Изд-во ТГУ, Тарту, 1982, стр. 80-82.
- [17] **Мальковский М.Г., Алексеев В.Д.** Семантическая память системы TULIPS-2. — В кн.: Человеко-машинные системы. Материалы семинара. — М.: Изд. МДНТП, 1977, стр. 29-36.
- [18] **Волкова И.А., Мальковский М.Г.** Представление знаний о языке общения с пользователем в системе TULIPS-2. — В кн.: IX Всесоюзный симпозиум по кибернетике. Тезисы симпозиума, М.: 1981, том 1, стр 160-163.
- [19] **Мальковский М.Г.** Семантические и прагматические предсказания при анализе текста. — В кн.: Семантика естественных и искусственных языков в специализированных системах. Тезисы. Л.: Изд-во ЛГУ, 1979, стр. 106-108.
- [20] **Пильщиков В.Н.** Язык программирования ПЛЭНЕР-БЭСМ. М.: Изд-во МГУ, 1978.
- [21] Автоматический перевод (сборник). — М.: Прогресс, 1971.
- [22] **Волкова И.А.** Лексическая адаптация в системе TULIPS-2. — В кн.: Диалог в автоматизированных системах, М.: Изд. МДНТП, 1981. стр. 76-81.
- [23] **Волкова И.А., Грацианова Т.Ю.** Обучение на лексическом уровне при диалоге с ЭВМ на естественном языке. — В кн.: Материалы XIX Всесоюзной студенческой

- научной конференции. Сер. Математика, Новосибирск, Изд-во НГУ, 1981, стр. 10–17.
- [24] **Волкова И.А.** Диалог с ЭВМ на естественном языке. Обучение на словарном уровне. — В кн.: Прикладная математика и математическое обеспечение ЭВМ, М.: Изд-во МГУ, 1981, стр. 109–111.
- [25] **Граудина Л.К., Ицкович В.А., Катлинская Л.П.** Грамматическая правильность русской речи. Опыт частно–стилистического словаря вариантов. — М.: Наука, 1976.
- [26] **Красиков Ю.В.** Теория речевых ошибок (на материале ошибок наборщика). — М.: Наука, 1980.
- [27] **Грис Д.** Конструирование компиляторов для цифровых вычислительных машин. — М.: Мир, 1975.
- [28] **Волкова И.А.** Диалог с ЭВМ на естественном языке. Адаптация и обучение на синтаксическом уровне. — Вестн. Моск. ун-та, Сер. Вычисл. матем. и киберн., 1982, №1, стр. 76-79.
- [29] **Винокурова Н.В., Волкова И.А., Грацианова Т.Ю., Мальковский М.Г.** Об одной модели русской морфологии. — В кн.: Актуальные проблемы прикладной математики и математического обеспечения ЭВМ, М.: Изд-во МГУ, 1982.

ПРИЛОЖЕНИЕ 1.

Условные сокращения:

И.	именительный падеж;
Р.	родительный падеж;
Д.	дательный падеж;
В.	винительный падеж;
Т.	творительный падеж;
П.	предложный падеж;
Ед.	единственное число;
Мн.	множественное число;
О.	одушевлённое;
Н.	неодушевлённое;
М.	мужской род;
Ж.	женский род;
С.	средний род;
НСВ	несовершенный вид;
СВ	совершенный вид;
наст.	настоящее время;
буд.	будущее время;
прош.	прошедшее время;
повелит.	повелительное наклонение;
деепр.	деепричастие;
1	первое лицо;
2	второе лицо;
3	третье лицо.

1.1. ТАБЛИЦА ФЛЕКСИЙ ПО П–КЛАССАМ СУЩЕСТВИТЕЛЬНЫХ МУЖСКОГО РОДА

П–класс		1	2	3	4	5	6	7	8	9
Примеры		завод артист	фонарь житель	чайник бульдог	марш хряц	месяц принц	случай герой	калий гений	южа- нин ар-	нож бо- гач
Ед.	И.	∅	ь	∅	∅	∅	и	и	∅	∅
	Р.	а	я	а	а	а	я	я	а	а
	Д.	у	ю	у	у	у	ю	ю	у	у
	В.	∅	ь	∅	∅	∅	и	и	–	∅
	Н									
	О	а	я	а	а	а	я	я	а	а
Мн.	Т.	ом	ем	ом	ем	ем	ем	ем	ом	ом
	П.	е	е	е	е	е	е	и	е	е
	И.	ы	и	и	и	ы	и	и	е	и
	Р.	ов	ей	ов	ей	ев	ев	ев	∅	ей
	Д.	ам	ям	ам	ам	ам	ям	ям	ам	ам
Т.	В.	ы	и	и	и	ы	и	и	–	и
	Н									
	О	ов	ей	ов	ей	ев	ев	ев	∅	ей
Т.	ами	ями	ами	ами	ами	ями	ями	ами	ами	

П–класс	1	2	3	4	5	6	7	8	9
Примеры	завод артист	фонарь житель	чайник бульдог	марш хрящ	месяц принц	случай герой	калий гений	южа- нин ар-	нож бо- гач
П.	ах	ях	ах	ах	ах	ях	ях	ах	ах

1.2. ТАБЛИЦА ФЛЕКСИЙ ПО П–КЛАССАМ СУЩЕСТВИТЕЛЬНЫХ ЖЕНСКОГО РОДА

П–класс	11	12	13	14	15	16	17	18	19
Примеры	карта коро-	неделя богиня	книга собака	туча кас-	улица волчи-	статуя фея	линия фурия	тет- радь	ночь мышь
Ед.	И. а	я	а	а	а	я	я	ь	ь
Р.	ы	и	и	и	ы	и	и	и	и
Д. П.	е	е	е	е	е	е	и	и	и
В.	у	ю	у	у	у	ю	го	ь	ь
Т.	ой/ою	ей/ею	ой/ою	ей/ею	ей/ею	ей/ею	ей/ею	ьо	ью
Мн	И. ы	и	и	и	ы	и	и	и	и
Р.	∅	ь	∅	∅	∅	й	и	ей	ей
Д.	ам	ям	ам	ам	ам	ям	ям	ям	ам
В. Н.	ы	и	и	и	ы	и	и	и	и
О.	∅	ь	∅	∅	∅	й	и	ей	ей
Т.	ами	ями	ами	ами	ами	ями	ями	ями	ами
П.	ах	ях	ах	ах	ах	ях	ях	ях	ах

П–класс		1	2	3	4	5	6	7	8	
Ед.	Крат- кая форма	М.	∅	∅	∅	∅	∅	и	∅	∅
		Ж.	а	я	а	а	а	я	а	а
Мн. (всех ро- дов)		С.	о	е	о	е	е	е	∅	о
		И.	ые	ие	ие	ие	ые	ие	ые	ие
		Р. П.	ых	их	их	их	ых	их	ых	их
		Д.	ым	им	им	им	ым	им	ым	им
		В. Н.	ые	ие	ие	ие	ые	ие	ые	ие
О.	ых	их	их	их	ых	их	ых	их		
Т.	ым	им	им	им	ым	им	ым	им		
Крат- кая		ы	и	и	и	ы	и	ы	и	

1.3. ТАБЛИЦА ФЛЕКСИЙ ПО П–КЛАССАМ СУЩЕСТВИТЕЛЬНЫХ СРЕДНЕГО РОДА

П–класс	21	23	25	26	27	28
Примеры	болото лицо	солнце чудо-	здание острие	горе поле	знамя имя	ущелье копье
Ед.	И. В. о	е	е	е	я	е
Р.	а	а	я	я	и	я
Д.	у	у	ю	ю	и	ю

П–класс		21	23	25	26	27	28
	Т. П.	ом е	ем е	ем и	ем е	ем и	ем е
Мн	И.	а	а	я	я	а	я
	Р.	∅	∅	и	ей	∅	и
	Д.	ам	ам	ям	ям	ам	ям
	В. О.	а	а	я	я	а	я
	Н.	∅	∅	–	–	–	–
Т. П.	ами ах	ами ах	ями ях	ями ях	ами ах	ями ях	

1.4. ТАБЛИЦА ФЛЕКСИЙ ПО П–КЛАССАМ СЛОВ АДЪЕКТИВНОГО СКЛОНЕНИЯ

П–класс			1	2	3	4	5	6	7	8	
Примеры			новый умный	синий лап-	редкий мягкий	свежий горя-	куцый узкопи-	длинно- шей тон-	смеш- ной	сухой лопо-	
Ед.	М. С.	И. И.	ни ое	ий ее	ий ое	ий ее	ый ее	ий ее	ой ое	ой ое	
	М. С.	Р. Д.	ого ому	его ему	ого ому	его ему	его ему	его ему	ого ому	ого ому	
	М.	В. Н. О.	ый ого	ий его	ий ого	ий его	ый его	ий его	ой ого	ой ого	
	С.	В.	ое	ее	ое	ее	ее	ее	ое	ое	
	М. С.	Т. П.	ым ом	им ем	им ом	им ем	ым ем	им ем	ым ом	им ом	
	Ж.	И. Р.Д. П.	ая ой	яя ей	ая ой	ая ей	ая ей	ая ей	ая ей	ая ой	ая ой
		В. Т.	ую ой/ою	юю ей/ею	ую ой/ею	ую ей/ею	ую ей/ею	ую ей/ею	ую ой/ою	ую ой/ою	

1.5. ТАБЛИЦА ФЛЕКСИЙ ПО П–КЛАССАМ СЛОВ МЕСТОИМЕННОГО СКЛОНЕНИЯ

П–класс			1	2
Примеры			дядин отцов	лисий мой
Ед.	М. С.	И. И.	∅ о	и е
	М. С.	Р. Д.	а у	его ему
	М.	В. Н. О.	∅ а	й его
	С.	В.	о	е
	М. С.	Т.	ым	им
	М. С.	П.	ом	ем
	Ж.	И. Р.Д.П.	а ой	я ей

П-класс		1	2
	В. Т.	у ой/ою	ю ей/ею
Мн.	И.	ы	й
	Р.П.	ых	их
	Д.	ым	им
	В. Н. О.	ы ых	и их
	Т.	ыми	ими

1.6. ТАБЛИЦА ФЛЕКСИЙ ПО П-КЛАССАМ ГЛАГОЛОВ

П-класс			1	2	3	4	5	6	7	8	9
Примеры			дуть ле-	гнуть вянуть	ло- вить	точить бежать	про- сить	рвать грызть	месты везти	зажечь испечь	тереть уме-
Инфинитив или НСВ буд.			ть	ть	ть	ть	ть	ть	та	ь	еть
НСВ наст. или СВ буд.	Ед.	1	ю	у	ю	у	у	у	у	у	у
		2	ешь	ешь	ишь	ишь	ишь	ешь	ешь	ешь	ешь
		3	ет	ет	ит	ит	ит	ет	ет	ет	ет
	Мн.	1	ем	ем	им	им	им	ем	ем	ем	ем
		2	ете	ете	ите	ите	ите	ете	ете	ете	ете
		3	ют	ут	ят	ат	ят	ут	ут	ут	ут
НСВ деепр. наст.			я	а/–	я	а	я	я	я	–	–
Повелит. НСВ и СВ	Ед.	2	й	и	и	и	и	и	и	и	и
	Мн.	2	йте	ите	ите	ите	ите	ите	ите	ите	ите
НСВ и СВ прош.	Ед.	м	л	л	л	л	л	л	л	∅	∅
		ж	ла	ла	ла	ла	ла	ла	ла	ла	ла
		с	ло	ло	ло	ло	ло	ло	ло	ло	ло
	Мн.		ли	ли	ли	ли	ли	ли	ли	ли	ли
Деепр.		в/вши	в/вши	в/вши	в/вши	в/вши	в/вши	в/вши	–/ши	–/ши	ев/ши

1.7. СПИСОК ЧЕРЕДОВАНИЙ

- | | |
|------------|-------------|
| 1. о → ∅ | 29. ч → к |
| 2. е → ∅ | 30. а → л |
| 3. е → й | 31. ова → у |
| 4. е → ь | 32. ева → у |
| 5. и → ∅ | 33. ева → ю |
| 6. и → ь | 34. ∅ → м |
| 7. л → ∅ | 35. а → ∅ |
| 8. ж → з | 36. я → ∅ |
| 9. за → ж | 37. у → ∅ |
| 10. ж → д | 38. ∅ → в |
| 11. ш → с | 39. ∅ → д |
| 12. са → ш | 40. ∅ → т |
| 13. ч → т | 41. ш → х |
| 14. та → ч | 42. ра → ер |
| 15. щ → т | |

16. та → щ
17. щ → ст
18. га → ж
19. ка → ч
20. ха → ш
21. ска → щ
22. да → ж
23. ста → щ
24. с → д
25. с → т
26. с → б
27. ж → г
28. ч → г

43. ы → о
44. е → о
45. а → м
46. ня → м
47. ы → уд
48. ня → йм
49. я → им
50. ∅ → ен
51. ла → ел
52. на → он
53. ва → ов
54. я → н
55. а → н
56. ф → ва

ПРИЛОЖЕНИЕ 2.

2.1. ТАБЛИЦА СООТВЕТСТВИЯ СЛОВАРНОЙ ИНФОРМАЦИИ ИЗ «ГРАММАТИЧЕСКОГО СЛОВАРЯ РУССКОГО ЯЗЫКА» А.А.ЗАЛИЗНЯКА СЛОВАРНОЙ ИНФОРМАЦИИ СИСТЕМЫ TULIPS–2.

TULIPS–2	Словарь А.А. Зализняка
Исключения	
МИ1	м(мо) — цифра и буква любые — с пометой ① не оканчивающиеся на «ь» и не слово «край»
МИ2	м(мо) — цифра и буква любые — с пометой ②; ж(жо) 2 * а — на «-ня».
МИ3	св(нсв) 7 — с пометой ⑨.
МИ4	св(нсв) 4 а — перед «-ить» — одиночная согласная;
МИ7	св(нсв) 5 а — глаголы без префикса «вы-» на «-ать», «-еть»; св(нсв) 6 а — глаголы без префикса «вы-» на «-ать»: перед «-ать» стоит л, н, р, с, з, т, д, к, г или х (с, з, т, д, к, г, х — одиночные согласные), а также слово «сыпать» и его производные; св(нсв) 7 а — глаголы без префикса «вы-»; св(нсв) 15 а;
МИ5	св(нсв) 3 а / 3° а — перед «-нуть» — гласная, а также слова «сты- св(нсв) 3° — перед «-нуть» — согласная: либо без пометы ⑤, либо с пометой [⑤]
МИ6	с(со) — цифра и буква любые — оканчивающиеся на «о» с пометой
МИ8	св(нсв) 3° а — перед «-нуть» — согласная либо без пометы ⑥ либо с пометой [⑥]; св(нсв) 7 — глаголы на «-зть».
МИ9	св(нсв) 5 — глаголы на «-ять»;
МИ10	св(нсв) 4 а — перед «-ить» — гласная.
МИ11	м(мо) — цифра и буква любые — с пометой ①, оканчивающиеся на «ь» и слово «край»
МИ12	ж(жо,мо-жо) — цифра и буква любые — с пометой ②.
МИ13	с (со) — цифра и буква любые — оканчивающиеся на «е» с пометой
МИ14	с(со) — цифра и буква любые — с пометой ①.
МИ15	св(нсв) 10;
МИ16	св(нсв) 6 — на «-ать»: перед «-ать» не стоит с,з,т,д,к,г,х,л,н,р.
МИ17	м 1(3,4,5) — с пометой П ₂ .
МИ18	м 2(6,7) — с пометой П ₂ .
МИ19	м 1(3,4,5) — с пометой Р ₀ .
МИ20	м 2(6,7) — с пометой Р ₂ .
МИ21	м 1(3,4,5) — с пометами Р ₂ и П ₂ .
МИ22	м 2(6,7) — с пометами Р ₂ и П ₂ .
МИ23	п<мс 1 а> — на «-ин», «-ын».
МИ24	п<мс I а> — на «-ин», «-ын».
МИ25	п 2 — без значка «*».
МИ26	п<мс 1 (буква любая)> — имена собственные
Чередования	
1 (∅)	м(мо) 1 * — перед последней согласной — «о»;
1 (ь)	м(мо) 2 * — —" — —" — —" —
1 (∅)	м(мо) 3 * — —" — —" — —" —
—" —	ж(жо) 1 * — перед последней согласной: к, г, х или шипящая (кроме 1 * d — перед последней согласной — шипящая);

TULIPS-2	Словарь А.А. Зализняка
<p>1 (ь) 1(∅) —" 1(ь) 1(∅) —" —" —" —" 1(ь) 1 (ть л ла ло ли в вши)</p>	<p>ж(жо) 2 к — перед последней согласной: к, г, х; ж(ясо) 3 * (кроме 3 * в) — перед последней согласной не стоит: ь, и, а(яо) 3 * в — перед последней согласной не стоит: ь, й. ж(яо) 8 *; с(со) 1 * — перед последней согласной: к, г, х; с(со) 1 * в; с(со) 3 * — перед последней согласной не стоит: ч, ш; п 1 * в — перед последней согласной — шипящая; п 3 * — перед последней согласной не стоит: ь, й и шипящая; Слово «тоlochь» и его производные; св(нсв) 10.</p>
<p>2 (∅) 2 (ь) 2 (∅) 2 (∅) —" —" 2 (ь) 2 (∅) —" —" —" —" —" —" —" —" —" —" —" 2 (ть л ла до ли в вши) 2 (еть ∅ ла ло ли ев ши)</p>	<p>м(мо) 1 * — перед последней согласной не стоит: «ле», «гласная+е»; м (мо) 2 * — —" —" —" м(мо) 3 * — перед последней согласной не стоит: «ле», «гласная+е», «согласная+ё» м(мо) 5 * — перед последней согласной не стоит: «ле», «гласная+е»; д(2о) 1 * — перед последней согласной не стоит: шипящая, к, г, х, ь, ж(жо) 1 * а — перед последней согласной — шипящая; ж(жо) 2 * — перед последней согласной не стоит: к, г, х, ь, й; ж(жо) 3 * а — перед последней согласной — шипящая; ж(жо) 5 *; с(со) 1 * — перед последней согласной не стоит: к, г, х, ь; с(со) 1 * в ; с(со) 3 * — перед последней согласной — ч или ш; с(со) 5 * — перед последней согласной не стоит ь (кроме слова «яй- п 1 * — перед последней согласной не стоит: ь, й (кроме 1 * в — пе- ред последней согласной шипящая); п 2 * — перед последней согласной не стоит: ь, й; п 3 * — перед последней согласной не стоит шипящая; Слово «жечь» и его производные; св(нсв) 5 — на «-еть»; св(нсв) 9.</p>
<p>3 (∅) —" —" 3 (∅) 3 (ь) 3 (∅) —" —" —"</p>	<p>м(мо) 1 * — перед последней согласной «гласная+е(ё)»; м(мо) 3 * — —" —" —" м(мо) 5 * — —" —" —" а(ао) 1 * — перед последней согласной стоит й; ж(жо) 2 * а — —" —" —" ж(жо) 3 * — —" —" —" п 1 * — —" —" —" п 2 * — —" —" —" п 3 * — —" —" —"</p>
<p>4 (∅) 4 (ь) —" —" 4 (∅) 4 (й)</p>	<p>м(мо) 1 * — перед последней согласной стоит «ле(лѐ)»; м(мо) 2 * — —" —" —" м(мо) 3 * — —" —" —" м(мо) 3 * в — перед последней согласной стоит: «ле(лѐ)», «соглас- м(мо) 5 * — перед последней согласной стоит «ле(лѐ)»; м(мо) 6 * ;</p>

TULIPS-2	Словарь А.А. Зализняка
4 (∅) —" —"	ж(жо) 1 * — перед последней согласной стоит «Ь»; ж(жо) 3 * — —" —" —"
4 (й) —" —"	ж(жо) 6 * В / 6 * d;
4 (∅) —" —"	с(со) 1 * — перед последней согласной стоит «Ь»; с(со) 5 * — —" —" —"
4 (й) —" —"	с(со) 6 * в; Слова «чей», «ничей»;
4 (∅) —" —"	п 1 * — перед последней согласной стоит «Ь»; п 2 * — —" —" —" п 3 * — —" —" —"
4 (и йте)	св(нсв) 11 / 11 *.
5 (ть л ла ло ли в вши)	св(нсв) 4.
6 (й) —" —"	ж(жо) 6 * а; с(со) 6 * а / 6 * d; п(числ.-п) <мс 6 * а>;
6 (ть л ла ло ли в вши)	св(нсв) 11 / 11 *.
7 (ю) —" —"	св(нсв) 4 — перед «-ить» стоит: п, б, м, в или ф; св(нсв) 5 — перед «-еть» стоит: п, б, м, в, ф, а также слово «спать» и его производные
8 (у) —" —"	св(нсв) 4 — на «-зить».
9 (ть л ла ло ли в вши)	св(нсв) 6 — на «-зать».
10 (у) —" —"	св(нсв) 4 — на «-дить»; св(нсв) 5 а/в — на «-деть».
11 (у) —" —"	св(нсв) 4 — на «-сить»; ев(нсв) 5 а/в — на «-сеть»;
12 (ть л ла ло ли в вши)	св(нсв) 6 — на «-сать».
13 (у) —" —"	св(нсв) 4 — на «-тить» (но не «-стить»); св(нсв) 5 а(в,с) — на «-теть» (но не «-стеть»).
14 (ть л ла ло ли в вши)	св(нсв) 6 — на «-тать» (но не «-стать»).
15 (у)	св(нсв) 4 а/в (-щ-) — на «-тить».
16 (ть л ла ло ли в вши)	св(нсв) 6 (-щ-) — на «-тать».
17 (у) —" —"	св(нсв) 4 а(в,с) — на «-стить»; св(нсв) 5 а(в,с) — на «-стеть».
18 (ть л ла ло ли в вши)	св(нсв) 6 — на «-чать».
19 (ть л ла ло ли в вши)	св(нсв) 6 — на «-кать».
20 (ть л ла ло ли в вши)	св(нсв) 6 — на «-хать».
21 (ть л ла ло ли в вши)	св(нсв) 6 — на «-скать».
22 (ть л ла ло ли в вши)	св(нов) 6 — на «-дать».

TULIPS-2	Словарь А.А. Зализняка
23 (ть л ла ло ли в вши)	св(нсв) 6 — на «-стать».
24 (ть)	св(нсв) 7 (-д-) — на «-сть», «-зть»
24 (ти)	св(нсв) 7 (-д-) — на «-сти», «-зти».
25 (та)	св(нсв) 7 (-т-).
26 (ти)	св(нсв) 7 (-б-).
27 (ель ет ей ете)	св(нсв) 8 (-г-), а также слова «жечь», «мочь», «лгать» и их произ-
28 (ь)	св(нсв) 8 (-г-), а также слова «жечь», «мочь» и их производные.
29 (ь ешь ет ем ете)	ев(нсв) 8 (-к-), а также слово «толочь» и его производные.
30 (ть л ла ло ли в вши)	св(нсв) 6 — на «-ать»: перед «-ать» стоит ц, б или м.
31 (ть л ла ло ли в ваш)	св(нсв) 2 — на «-овать».
32 (ть л ла ло ли в вши)	св(нсв) 2 — на «-евать»: перед «-евать» — шипящая.
33 (ть д ла ло ли в вши)	св(нсв) 2 — на «-евать»: перед «-евать» — не шипящая.
34 (ть л ла ло ли в вши)	св(нсв) 15;
34 (∅)	п 1 — на «-нный» с пометой ①;
34 (∅ а о ы)	п 1 — на «-нный» с пометой ②;
34 (0 ши ла ло ли)	св(нсв) 3°.
35 (ть л ла ло ли в вши)	св(нсв) 5, 6, 6°, 13 — на «-ать».
36 (ть л ла ло ли в вши)	св(нсв) 5, 6, 14 — на «-ять».
37 (ть л ла ло ли в вши)	св(нсв) 3, 3°.
38 (ть л ла ло ли в вши)	св(нсв) 16;
38 (ю ешь ет ем ете зат)	св(нсв) 13.
39 (л ла ло ли)	св(нсв) 7 (-д-).
40 (л ла до ли)	св(нсв) 7 (-т-).
41 (е)	п 3 — на «-хий».
42 (ть л ла ло ли в вши)	св(нсв) 6° — на «-братъ», «-драть».
43 (ть л ла ло ли в вши)	ев(нов) 12 — на «-ыть».
44 (ть л ла ло ли в вши)	св(нсв) 12 — на «-петь».
45 (ть л ла ло ли в вши)	св(нсв) 14 / 14 * (-м-) — на «-ать».
46 (ть л ла ло ли в вши)	св(нсв) 14 / 14 * — на «-принять».
47 (ть л ла ло ли в вши)	св(нсв) — на «-быть».

TULIPS-2		Словарь А.А. Зализняка
48 (ть л ла ло ли в вши)		св(нсв) 14 / 14 * — на «-занять», а также слова «нанять», «перенять», «донять», «пронять», «унять».
49 (ть л ла ло ли в вши)		св(нсв) 14 / 14 * — «обнять», «поднять», «снять», «отнять» и их производные.
50 (я)		с 8° — на «мя».
51 (ть л ла ло ли в вши)		св(нсв) 6 — на «-стлать».
52 (ть л ла ло ли в вши)		св(нсв) 5 — на «-гнать».
53 (ть л ла ло ли в вши)		св(нсв) 6° — на «-звать».
54 (ть л ла ло ли в вши)		св(нсв) 14 / 14 * (-н-) — на «-ять».
55 (ть л ла ло ли в вши)		св(нсв) 14 / 14 * (-н-) — на «-ать».
56 (ю, евш, ет, ете, ют, ем)		св (нсв) 13.
М-	П-	
5	1	п<мс 1 (буква и цифра любые)> — на «-ин», «-ын», «-ов», «-ев».
	2	п(числ.-п) <мс-п б* (буква любая)>, а также слова «мой», «твой», «чей», «ничей», «кой», «ником».
7	1	м(мо) 1 (с любой буквой) и 5 в.
	2	м(мо) 2 (с любой буквой).
	3	м(мо) 3 (с любой буквой).
	4	м(мо) 4 а (с, d, e, f).
	5	м(ью) 5 а (с, d, e, f).
	6	к(ью) 6 (с любой буквой).
	7	м(мо) 7 (с любой буквой).
	8	мо 1°.
	9	м(мо) 4 в.
	11	ж(жо,мо-жо) 1 (с любой буквой) и 5 в.
	12	—"— 2 (с любой буквой).
	13	—"— 3 (с любой буквой).
	14	—"— 4 (с любой буквой).
	15	5 а (с, d, e, f).
	16	—"— 6 (с любой буквой).
	17	—"— 7 (с любой буквой).
	18	—"— 8 (основа не на шипящую).
	19	—"— 8 (основа на шипящую).
	21	с(со) 1 (с любой буквой). 3а, 3с, 4в, 4d, 4f, 5в, 5d, 5f.
	23	с(со) 4а, 4с, 4е, 5а, 5с, 5е.
	25	с(со) 7 (с любой буквой).
	26	с(со) 2 (с любой буквой).
	27	с(со) 8 (с любой буквой).
	28	с(со) 6 * (с любой буквой).
8	1	п 1 а(с, d, e, f).
	2	п 2 (с любой буквой).
	3	п 3 а(с, d, e, f).
	4	п 4 а(с, d, e, f).
	5	п 5 (с любой буквой).

TULIPS–2		Словарь А.А. Зализняка
	6	п 6 (с любой буквой).
	7	п 1 в.
	8	п 3 в / 4 в.
9	1	св(нсв) 1 (2, 6, 10, 11, 12, 13); 6 — перед «-ать» не стоит: с, з, т, д, к, г, х и не слово «гнать» и 6 — на «-ать»
	2	св(нсв) 3(3°,6); 6 — перед «-ать» стоит: с, з, т, д, к, г, х, и слово
	3	св(нсв) 4 — перед «-ить» стоит: л, н, р, п, б, м, в, ф или гласная;
	4	св(нсв) 5 — на «-еть» «-ять»: перед «-еть» стоит: п, н, р, п, б, м, в, и св(нсв) 4 — перед «-ить» стоит: ж, ч, ш, щ; св(нсв) 5 — на «-ать», и не слово «гнать»
	5	св(нсв) 4(5) — не принадлежащие П-классам 3 и 4.
	6	св(нсв) 6° (7, 14, 15, 16); 7 — на «-еть», «-зть».
	7	св(нсв) 7 — на «-сти», «-эти».
	8	св(нсв) 8.
	9	св(нев) 9.

2.2. ПРАВИЛА РАБОТЫ С ПРОЦЕДУРОЙ АВТОМАТИЗИРОВАННОГО ПОПОЛНЕНИЯ СЛОВАРЯ СИСТЕМЫ TULIPS–2 ПО ДАННЫМ ИЗ «ГРАММАТИЧЕСКОГО СЛОВАРЯ РУССКОГО ЯЗЫКА» А.А. ЗАЛИЗНЯКА

Как уже отмечалось в основном тексте, непосредственно обращается к процедуре автоматизированного пополнения словаря системы TULIPS–2 КОД и записывает результат её работы в словарь процедура 333. Чтобы записать в словарь системы новые слова, пользователь должен вызвать процедуру 333 и, получив от неё приглашение к началу работы, последовательно набирать на клавиатуре терминала слова и соответствующую им словарную информацию из словаря А.А. Зализняка по следующим правилам:

- 1) значок «°» обозначается " ;
- 2) значок «*» обозначается * ;
- 3) цифра в кружке записывается как цифра в круглых скобках;
- 4) П₂ (P₂) обозначается П2 (P2);
- 5) символы в угловых скобках записываются как символы в круглых скобках;
- 6) буква в круглых скобках и в тире: (—) записывается как буква в круглых скобках и в знаках равенства: (=);
- 7) значки X и \boxed{X} обозначаются + ;
- 8) значок ~ обозначается — ;
- 9) словарная статья слова заключается в круглые скобки, слово и словарная статья заключаются в общие круглые скобки;
- 10) все элементы словарной статьи пишутся через пробел, через пробел пишутся также значки * (1 *) и « (3 «);
- 11) вместо ссылок на аналогичные словарные статьи у других слов пишутся сами словарные статьи;
- 12) слова в форме множественного числа, у которых словарная статья начинается с «мн. от ...» пишутся в форме единственного числа со стоящими после предлога «от» элементами словарной статьи. Например, если надо ввести из словаря А.А. Зализняка слово «лёгкие» со словарной статьёй «мн. от с<п 3 а>», то надо набрать на терминале следующие данные:

ЛЕГКОЕ (С (П 3А)) ;

- 13) словесные пояснения в круглых скобках опускаются;
- 14) словесные пояснения в квадратных скобках опускаются;

- 15) символы, стоящие в квадратных скобках (например, [⊕]), либо опускаются, либо пишутся без квадратных скобок;
- 16) предлоги «в» и «на», стоящие в круглых скобках после символов П₂ или Р₂, опускаются;
- 17) знак «§» с рядом стоящей цифрой опускаются;
- 18) пометы «ё» или «о» опускаются;
- 19) значок «●» и стоящие после него символы опускаются;
- 20) знак «:» и стоящие за ним слова опускаются;
- 21) знак «→» опускается;
- 22) знак «/» и стоящая после него буква опускаются;
- 23) между основным (первым) буквенным символом и цифрой запятая не ставится;
- 24) значок «◊» и всё, что стоит после него до первого разделительного знака (т.е. до «Δ», «◊», «●», «//», «'», «:», «;», «•») и скобок, охватывающих данный знак) справа, опускаются;
- 25) значок «//» используется для записи вариантов, его действие распространяется от ближайшего разделительного знака слева до ближайшего разделительного знака справа. При наличии значка «//» надо вводить все варианты данного слова или его словарной статьи отдельно и уже без «//». Например, чтобы ввести слово «шампунь» со словарной статьёй «м 2 а // ж 8 а» надо сначала ввести вариант шампунь (м 2 а) ,

а затем

шамшунь (ж 8 а) ,

в для слова «бактерия» (словарная статья — «ж // жо, 7 а») надо ввести такие данные:

бактерия (ж 7 а) и

бактерия (жо 7 а) .

Для слова «закалённый» со словарной статьёй вида п 1* а/в ⊕, ё [// п 1* а ⊕, §8] записываются либо два варианта словарной статьи:

закалённый (п 1 * а ⊕) и

закалённый (п 1 * а ⊕) ,

либо только первый;

- 26) если среди символов словарной статьи встречается знак «;» (их может быть несколько), то слово надо вводить таким образом: а) если после «;» (до следующего разделительного знака) стоит одиночный основной буквенный символ (м, мо, ж, жо, с, со или мо–жо), то сначала надо ввести слово со словарной статьёй до «;», а затем со словарной статьёй, где первым стоит один из указанных выше основных буквенных символов, а далее записывается словарная статья, стоящая до «;», заключённая в круглые скобки; б) если после «;» стоит словарная статья, то слово вводится по очереди с каждой из словарных статей. Например, чтобы записать в словарь системы слово «штатский» со словарной статьёй «п 3 а X ~ ; мо», надо ввести следующие данные;

(штатский (п 3 а + -)) и

(штатский (мо(п 3 а + -)));

а для слова «кукушка» со словарной статьёй «жо 3* а (птица); ж 3* а (паровоз)» — такие данные:

(кукушка (жо 3 * а))

и

(кукушка (ж 3 * а));

27) если среди символов словарной статьи встречается значок «Δ» (признак различных особенностей словоизменения), то в общем случае такое слово надо вводить вручную. Но в следующих наиболее часто встречающихся и простых случаях еловл с символом «Δ» можно записывать в словарь системы с помощью процедуры 333 (во всех этих случаях символ «Δ» опускается):

а) после «Δ» стоит «мн. основа /-ья, -ьев, -ьям»;

б) после «Δ» стоит «Р. мн. <словоформа>»;

в) после «Δ» стоит «сравн. <словоформа>», «исх. форма нерегул.:...» или «прич. страд....»;

г) «Δ» стоит в словарных статьях глаголов на «-звать», «-толочь», «-гнать», «-братъ», «-драть», «-петь», «-мереть», «-хотеть», «-быть» (или они же с частицами «-ся»), а также слов «житьё», «питьё», «бытие», «небытие», «гражданин», «кружево», «зло», «озеро», «утро», «хлопоты», «гумно», «ляпсить», «стоять», «судия», «паремия», «лития», «знамя», «папайя», «пирайя», «секвойя», «паранойя», «аллилуйя», «начаться».

ПРИЛОЖЕНИЕ 3.

3.1. ТАБЛИЦА ФЛЕКСИЙ ДЛЯ ПРОЦЕДУР СИНТЕЗА

С–класс	П–класс	Флексии
5	1	((() а у (а ())) ым ом) (а ой ой у ой ой) (о а у о ш ом) (ы ых ым (ых ы) ыми ых))
	2	((и его ему (его и) им ем) (я ей ей ю ей ей) (е его ему е им ем) (и их им (их и) ими их))
6	1	((() а у (а ())) ом е) (ы ов ам (ов ы) ами ах))
	2	((ь я ю (я ь) ем е) (и ей ям (ей и) ями ях))
	3	((() а у (а ())) ом е) (и ов ам (ов и) ами ах))
	4	((() а у (а ())) ем е) (и ей ам (ей и) ами ах))
	5	((() а у (а ())) ем е) (ы ев ам (ев ы) ами ах))
	6	((й я ю (я и) ем е) (и ев ям (ев и) ями ях))
	7	((й я ю (я и) ем и) (и ев ям (ев и) ями ях))
	8	((() а у (а ())) ом е) (е () ам (()) ами ах))
	9	((() а у (а ())) ом е) (и ей ам (ей и) ами ах))
	11	((а ы е у ой е)(ы () ам (() ы) ами ах))
	7	12
13		((а и е у ой е)(и () ам: (() и) ами ах))
14		((а и е у ей е)(и () ам (() и) ами ах))
15		((а ы е у ей е)(ы () ам (() ы) ами ах))
16		((я и е ю ей е)(и и ям (и и) ями ях))
17		((я и и ю ей и)(и и ям (и и) ями ях))
18		((ь и и ь ю и)(и ей ям (ей и) ями ях))
19		((ь и з ь ю п)(и ей ам (ей и) ами ах))
21		((о а у о ом е) (а () ам (() а) аш ах))
23		((е а у е ем е)(а () ам (() а) ами ах))
25		((е я ю е ем и) (я и ям ями ях))
26		((е я ю е ем е)(я ей ям ями ях))
27		((я и и я ем и) (а () ам ами ах))
28		((е я ю е ем е)(я и ям ями ях))
8	1	((ый ого ому (ого ый) ым ом) (ая ой ой ую ой ой) (ое ого ому ое ым ом) (ые ых ым (ых ые) ыми ых))
	2	((ий его ему (его ий) им ем) (яя ей ей юю ей ей) (ее его ему ее им ем) (не их им (их ие) ими их))
	3	((ий ого ому (ого ий) им ом) (ая ой ой ую ой ой) (ое ого ому ое им ом) (ие их им (их ие) ими их))
	4	((ий его ому (его ий) им ем) (ая ей ей ую ей ей) (ее его ему ее ым ем) (ие их игл (их ие) ими их))
	2	3
17	7	ее
	8	е
9	1	((()ю ем)(ешь ете)(ет ют)) (л ла ло ли)(й йте) ть)
	2	((()у ем)(ешь ете)(ет ут)) (л ла ло ли)(и ите) ть)
	3	((()ю им)(ишь ите)(ти ят)) (л ла ло ли)(и ите) ть)
	4	((()у им)(ишь ите)(ит ат)) (л ла ло ли)(и ите) ть)
	5	((()у им)(ишь ите)(ит ят)) (л ла ло ли)(и ите) ть)
	6	((()у ем) (ешь ете)(ет ут)) (л ла ло ли)(и ите) ть)

	7	(((у ем)(ешь ете)(ет ут)) (л ла ло ли)(и ите) ти)
	8	(((у ем)(ешь ете)(ет ут)) (()) ла ло ли)(и ите) ь)
	9	(((у ем)(ешь ете)(ет ут)) (()) ла ло ли)(и ите) еть)
19	1	(я () в)
	2	(а () в)
	3	(я () в)
	4	(а () в)
	5	(я () в)
	6	(я () в)
	7	(я () ши)
	8	(и () ши)
	9	(* () ев)

3.2. ТАБЛИЦА ЧЕРЕДОВАНИЙ ДЛЯ ПРОЦЕДУР СИНТЕЗА

Номер чередования	Чередование
1	М–клас = 9, П–клас = 1, тогда (<основа>о), иначе (<основа без 1ПБ*> о <1ПБ>)
2	М–клас = 9, П–клас Э (3,5), тогда (<основа>е), иначе (<основа без 1ПБ>е<1ПБ>)
3	(<основа без 2ПБ>е<1ПБ>)
4	М–клас ∈ (5,9) или М–клас = 7 и П–клас в ∈ (6,16,28), тогда (<основа без 1ПБ>е), иначе (<основа без 2ПБ>е<1ПБ>)
5	(<основа>и)
6	(<основа без 1ПБ>и)
7	(<основа>л)
8	(<основа без 1ПБ>ж)
9	(<основа без 1ПБ>за)
10	(<основа без 1ПБ>ж)
11	(<основа без 1ПБ>ш)
12	(<основа без 1ПБ>са)
13	(<основа без 1ПБ>ч)
14	(<основа без 1ПБ>га)
15	(<основа без 1ПБ>щ)
16	(<основа без 1ПБ>га)
17	(<основа без 2ПБ>щ)
18	(<основа без 1ПБ>га)
19	(<основа без 1ПБ>ка)
20	(<основа без 1ПБ>ха)
21	(<основа без 1ПБ>ска)
22	(<основа без 1ПБ>да)
23	(<основа без 1ПБ>ста)
24	(<основа без 1ПБ>с)
25	(<основа без 1ПБ>с)
26	(<основа без 1ПБ>с)
27	(<основа без 1ПБ>ж)
28	(<основа без 1ПБ>ч)
29	(<основа без 1ПБ>ч)
30	(<основа без 1ПБ>а)

Номер чередования	Чередование
31	(<основа без 1ПБ>ова)
32	(<основа без 1ПБ>ева)
33	(<основа без 1ПБ>ева)
34	(<основа без 1ПБ>)
35	(<основа>а)
36	(<основа>я)
37	(<основа>у)
38	(<основа без 1ПБ>)
39	(<основа без 1ПБ>)
40	(<основа без 1ПБ>)
41	(<основа без 1ПБ> ш)
42	(<основа без 2ПБ> ра)
43	(<основа без 1ПБ>ы)
44	(<основа без 1ПБ>е)
45	(<основа без 1ПБ>а)
46	(<основа без 1ПБ>ня)
47	(< основа без 1ПБ>ы)
48	(<основа без 1ПБ>ня)
49	(<основа без 1ПБ>я)
5С	(<основа без 2ПБ>)
51	(<основа без 2ПБ>ла)
52	(<основа без 2ПБ>на)
53	(<основа без 2ПБ>ва)
54	(<основа без 1ПБ>я)
55	(<основа без 1ПБ>а)
56	(<основа без 2ПБ>)

3.3. ТАБЛИЦА ИСКЛЮЧЕНИЙ ДЛЯ ПРОЦЕДУР СИНТЕЗА

Тип исключения	Исключение
МИ1	((2 1)(2 (1 4)) а)
МИ2	((2 (24))(2 2)0)
МИ3	((3) я)
МИ4	((5 2 0 1) ь)
МИ5	((3 0 1 1)())
МИ6	((2 (2 4)) (2 2) ов)
МИ7	((5 2 0 2) ьте)
МИ8	((3) ши)
МИ9	((5 2 0 1) й)
МИ10	((5 2 0 2) й т е)
МИ11	((2 1)(2 (1 4)) я)
МИ12	((2 (2 4))(2 2) ей)
МИ13	((2 (2 4))(2 2) ев)
МИ14	((2 1)(2 (1 4)) и)
МИ15	((5 2 0 1) и)
МИ16	((5 2 0 2) ите)
МИ17	((1 8)(1 8) у)

Тип исключения	Исключене
МИ18	((1 8)(1 8) ю)
МИ19	((1 7)(1 7) у)
МИ20	((1 7) (1 7) ю)
МИ21	((1 (7 8))(1 (7 8)) у)
МИ22	((1 (7 8))(1 (7 8)) ю)
МИ23	((0 (1 3) 1 3)(0 (1 3) 1 3) ому)
МИ24	((1 (1 3) 1 (2 4))(2 (1 3) 1 2) ого)
МИ25	((1 1) ь)
МИ26	((1 1 1 6)(2 (1 3) 1 6) е)