

Санкт-Петербургский государственный университет

Филологический факультет

Кафедра общего языкознания

**Психолингвистическое исследование структуры
ментального лексикона на материале русских глаголов**

Дипломная работа студентки V курса
отделения теории языкознания
Слюсарь Наталии Анатольевны

Научный руководитель:
д.ф.н., д.б.н., проф. Т. В. Черниговская

Санкт-Петербург

2003

I. Современные теории об устройстве ментального лексикона.	4
1. Односистемный и двусистемный подходы.....	5
2. Диспут о прошедшем времени.....	10
3. Основные проблемы модели "Слова и Правило".....	15
II. Проблема существования в ментальном лексиконе стандартных правил.....	19
1. Эксперимент со взрослыми носителями: дизайн.....	19
2. Эксперимент со взрослыми носителями: результаты.....	24
3. Эксперименты с детьми: сравнение результатов.....	30
III. Проблема соотношения конструируемых и анализируемых форм и форм, хранящихся в памяти.	36
1. Классический эксперимент Алегре и Гордона.....	36
2. Почему невозможно повторить эксперимент Алегре и Гордона с русскими глаголами?.....	37
3. Подготовка нового эксперимента на материале русского языка.....	38
4. Экспериментальный дизайн.....	42
5. Статистическая обработка данных.....	47
6. Результаты эксперимента.....	52
IV. База данных по русскому глаголу на основе "Грамматического словаря русского языка".....	57
1. Описание базы данных.....	57
2. Сведения о глагольной системе русского языка, полученные при помощи базы данных.....	62
3. Сведения об акцентных парадигмах.....	66
4. Заключение.....	71
V. Библиография.....	73
VI. Приложения.....	79
Приложение 1. Одноосновная Система (OSM).....	79
Приложение 2. Соотношение Академической Грамматики и Одноосновной Системы (OSM).....	83
Приложение 3. Соотношение системы глагольных классов, принятых в словаре Зализняка, и Одноосновной Системы (OSM).....	86

Приложение 4. Статистические методы обработки данных.	89
Приложение 5. Значения условных помет и список акцентных парадигм в базе данных.	96
Приложение 6. Списки беспроставочных и немотивированных глаголов разных классов.	100

I. Современные теории об устройстве ментального лексикона.

Этот диплом представляет собой итог четырехлетней работы в рамках российско-американского проекта, цель которого – проверить некоторые современные гипотезы об устройстве ментального лексикона на материале русской глагольной морфологии. Российской частью проекта руководит профессор Санкт-Петербургского Государственного университета Т.В. Черниговская, а американской – профессор К. Гор из Мэрилендского университета в США. Работа выполнена при поддержке РГНФ (грант № 00-04-00338a) и РФФИ (грант № 00-15-9885). Результаты части исследований, проведенных в рамках нашего проекта, были опубликованы (Черниговская и др. 1999; Chernigovskaya et al. 2000a; Chernigovskaya et al. 2000b; Слюсарь 2000; Gor et al. 2001; Слюсарь в печати).

Дать точное определение понятия "ментальный лексикон" невозможно хотя бы потому, что многими учеными оно понимается по-разному. Однако все сходятся на интуитивном представлении о том, что ментальный лексикон – это, по сути, словарь родного языка, который хранится у нас в голове. Существование такого словаря позволяет нам узнавать знакомые слова в звучащем и письменном тексте и понимать их смысл и, наоборот, подыскивать для передачи своей мысли нужные слова и озвучивать или записывать их. Как и обыкновенный словарь, ментальный лексикон должен быть снабжен для этого системами кодирования, с одной стороны, звуковых и графических образов и с другой – смыслов. Чтобы было возможным не только хранение, но и поиск, планы выражения и планы содержания входящих в него единиц должны быть систематизированы и он должен быть снабжен различными поисковыми системами. Говоря на родном языке, мы употребляем правильные грамматические формы, а слушая или читая, правильно их интерпретируем. Следовательно, в ментальном лексиконе существуют механизмы, которые при порождении позволяют находить нужные формы слов или создавать их по определенным правилам, а при восприятии – анализировать их. Кроме того, ментальный лексикон постоянно пополняется. Обычно, услышав новое слово один или несколько раз, человек начинает свободно им пользоваться, то есть узнает на

письме и на слух, произносит, понимает и употребляет его сам в новом контексте и в новых синтаксических конструкциях, в новых формах.

Таким образом, можно сказать, что любое исследование из области психолингвистики так или иначе относится к изучению ментального лексикона. Очертим более узкий круг проблем, которыми занимается наш проект. Нас интересует грамматическая составляющая ментального лексикона. Как именно человек усваивает, хранит, порождает и воспринимает грамматические формы слов? Заметим, что, хотя экспериментаторы могут исследовать аспекты какого-то одного из перечисленных выше вопросов, полноценная теория в этой области должна ответить на все. К сожалению, в современной психолингвистике этого пока не произошло. Существует ряд проблем, которые раскалывают ученых на два лагеря.

1. Односистемный и двусистемный подходы.

Пожалуй, главное противопоставление в этой области психолингвистики на данном этапе – это противопоставление так называемых односистемного и двусистемного подходов. Первый из них был разработан в рамках коннекционизма и других вариантов сетевого представления морфологии. Создатели двусистемного подхода в большинстве своем являются сторонниками модулярной теории и символических правил.

Сторонники двусистемного подхода делят все формы на правильные и неправильные. Первые образуются при помощи символического правила, которое прибавляет к основе нужный аффикс (или, например, инфикс – в зависимости от языка). Вторые хранятся целиком в ассоциативной памяти. Считается, что правильные и неправильные формы относятся к разным подмодулям внутри языкового модуля. Понятие "модуль" стало в современной науке совершенно неоднозначным и применение его в данной ситуации несколько спорно. Имеется в виду, что правильные и неправильные формы хранятся в разных участках мозга и обрабатываются посредством разных, никак не связанных друг с другом механизмов. Эти подмодули являются врожденными, то есть человек генетически запрограммирован на то, чтобы искать в словоизменительной морфологии "стандартное правило", а найдя его, разделить все формы на правильные и неправильные.

В принципе, двусистемный подход можно считать продолжением традиционного представления о том, что для одних форм мы запоминаем правила, а другие заучиваем списками. Первая попытка предложить детально разработанную альтернативную теорию была сделана коннекционистами. В 1986 году вышла эпохальная книга "Parallel Distributed Processing" (Rumelhart and McClelland 1986), в которой были сформулированы основополагающие идеи коннекционизма. Во-первых, коннекционисты выступили против каких бы то ни было врожденных модулей. С их точки зрения, мозг изначально является единой нейронной сетью, и участки, ответственные за выполнение различных заданий, формируются в нем в результате постоянной переработки входящей информации и обратной связи с окружающей действительностью. Иначе говоря, в отличие от модулярной теории, основной акцент ставится на обучение, а не на врожденные механизмы (хотя оба начала, безусловно, играют важную роль и в том, и в другом подходе). Первоначально в нейронной сети "все связано со всем", и, хотя некоторые связи со временем слабеют, во многих случаях, когда символисты настаивают на нескольких автономных процессах, коннекционисты постулируют один или несколько взаимосвязанных. Кроме того, коннекционисты считают, что в мозгу не существует символических правил. Процессы, традиционно представлявшие человеку в виде применения определенных правил, могут осуществляться системой, в которой нет представления о правилах.

Все эти утверждения были проиллюстрированы на примере моделей нейронных сетей, которые обучались выполнять те или иные задания. Одна из этих моделей осваивала формы Прош. вр. от английских глаголов. Правильные и неправильные глаголы обрабатывались при помощи одних и тех же алгоритмов в рамках одной и той же сети. Коннекционисты хотели показать, что только за счет того, что формы Прош. вр. с суффиксом *-ed* частотнее и этот суффикс не зависит от фонологического окружения¹, новые слова попадают именно в этот класс и он вообще воспринимается как "правильный".

Кроме коннекционистов сетевой подход к морфологии разрабатывается Джоан Байби (Bybee 1985, 1988, 1995), когнитивистами в лице Рональда Лангакера (Langacker

¹ Например, чтобы образовать Прош. вр. при помощи аблаута, надо иметь соответствующие гласные в корне, а *-ed* можно прибавить к любой основе.

1987, 1988) и их последователями. Между этими школами существуют некоторые различия, однако их разъяснение потребовало бы углубления в сложнейшие "технические" детали, связанные с устройством нейронных сетей. Это показалось нам нецелесообразным, тем более, что разногласия не фундаментальны. В остальном же отличие Байби и Лангакера от коннекционистов носит, как нам кажется, скорее методологический, чем теоретический характер. Байби и Лангакер строят свои модели "на бумаге", формулируя общие принципы и иллюстрируя их на примере разных языков. Коннекционисты же в большинстве случаев работают с моделями нейронных сетей и, соответственно, детально изучают какой-нибудь конкретный процесс на материале конкретного языка. Кроме того, их интересует и большой круг лингвистических проблем: например, есть сети, которые обучаются закону сохранения материи (точнее, представлению о том, что предметы непрерывны во времени, а не исчезают каждый раз, когда скрываются из вида)² и заданию "сбалансируй бусины"³.

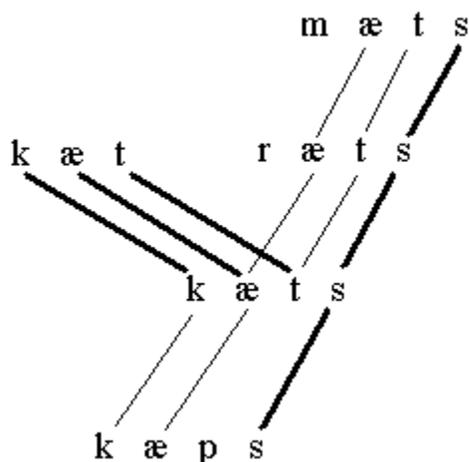
Возможности таких моделей пока очень ограничены, и поэтому не существует сети, которая могла бы не то что целиком освоить какой-то язык или его морфологию, а даже какую-то ее малую часть, скажем, все формы английского глагола. Классическая модель, описанная в книге "Parallel Distributed Processing" учится получать в качестве стимулов формы Наст. вр. от английских глаголов и выдавать в качестве реакций формы Прош. вр. Ни о значении этих форм (лексическом и грамматическом), ни о том, как их можно использовать, речи не идет. Стимулы даются в символической записи, которая напоминает фонологическую транскрипцию и встроена в систему изначально (то есть все проблемы, связанные с фонетикой и фонологией, также снимаются).

В связи с этим мы решили проиллюстрировать сущность сетевого подхода примерами из статьи Байби (Bybee 1995): по ним легче понять, как язык в общем и целом,

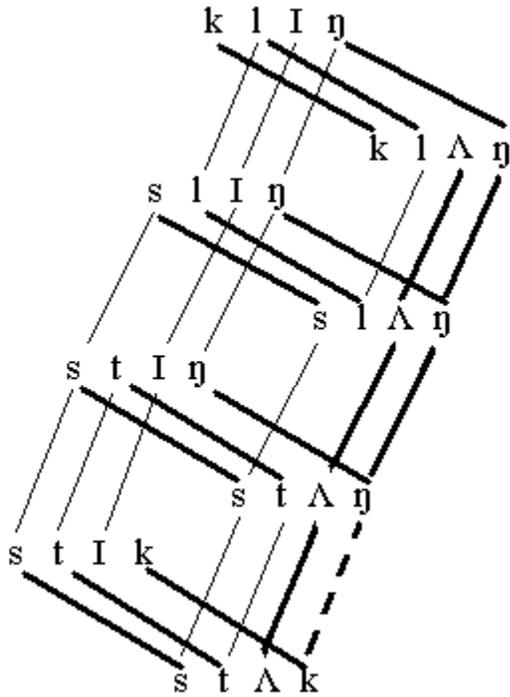
² Коннекционисты считают себя преемниками Пиаже, который первым выдвинул идею о том, что это представление – благоприобретенное.

³ Задание в общих чертах таково: имеется виртуальная палочка с центром равновесия и пятью ответствиями справа и слева от него. В некоторые из этих отверстий могут быть положены бусины. Понятно, что чем дальше бусина от центра, тем сильнее она наклоняет палочку, то есть тем больше ее относительный вес. Программа учится считать суммарный относительный вес справа и слева от центра равновесия при любом расположении бусин.

а не какая-то его конкретная маленькая часть, может быть представлен в ассоциативной сети. Вот схема, в общих чертах показывающая, как хранятся в такой сети регулярные формы Мн. ч. существительных (*cat – cats* "кот" – "коты", выше на рисунке созвучные слова *rats* "крысы", *mats* "циновки", ниже – *caps* "кепки"). Так как предполагается, что графическая форма слов вторична и привязана к звуковой, слова условно записаны в фонологической транскрипции.



Как мы видим из этой таблицы, все формы хранятся целиком. Аффикс *-s* нигде не представлен сам по себе, он существует только в составе форм Мн. ч. Слова существуют в виде дистрибутивных, то есть расчлененных, представлений. В данном случае это, по сути, означает, что их план выражения состоит из относительно автономных звуков, а план содержания – из автономных смысловых единиц. Между одинаковыми звуками и смыслами образуются связи. "Материальная база" лексического значения "кот" – связи, соединяющие все формы этого слова, как это показано на рисунке. "Материальная база" грамматического значения Мн. ч., содержащегося в суффиксе *-s*, – связи, соединяющие все суффиксы *-s*. Таким образом, не членив формы, мы достигаем того же эффекта, что и традиционное морфологическое членение. Для сравнения приведем схему ассоциативной сети, в которой хранятся формы Прош. вр. от группы английских неправильных глаголов (*cling – clung* "прилипнуть"; *sling – slung* "бросать, раскручивая"; *sting – stung* "ранить"; *stick – stuck* "ткнуть, засунуть"). Как мы видим, здесь использованы те же принципы, что и в случае с регулярным Мн. ч.



После выхода в свет книги Руммельхарта и МакКлеланда сторонники модулярной теории подвергли ее шквальной критике. Особенно обсуждалась главная лингвистическая модель – сеть, обучавшаяся образовывать формы Прош. вр. Вскоре появились многочисленные публикации, в которых символисты указали на серьезные недостатки этой модели и предложили свою, которая, с их точки зрения, лучше объясняла многие факты естественного языка. Именно тогда символисты детально разработали двусистемный подход. Мы решили привести здесь последнюю, "исправленную и дополненную" версию этой модели. Она называется *Words and Rule* ("Слова и Правило"), и была разработана Стивенем Пинкером и его единомышленниками (Pinker 1995, 1999). Вот ее схематическое изображение:



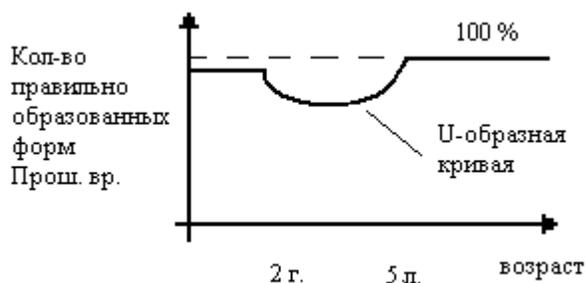
Механизм прост: для образования формы Прош. вр. к основе глагола применяется стандартное правило (для английского языка это прибавление суффикса *-ed*) в том случае, если не подействует блокирующее правило. Сущность блокирующего правила заключается в том, что оно останавливает применение стандартного правила, если форма Прош. вр. данного глагола содержится в лексиконе. В лексиконе хранятся формы Прош. вр. от всех неправильных глаголов. Стандартное правило также называется дефолтом (default). Итак, название модели отражает два ее компонента, два модуля: слова, хранящиеся в лексиконе, и стандартное правило, при помощи которого образуются формы от всех остальных слов.

2. Диспут о прошедшем времени.

Полемика между символистами и сторонниками сетевой теории длится до сих пор и получила название "Диспута о прошедшем времени". Вот его краткая хронология. Все началось с коннекционистской сети, образывавшей формы прошедшего времени от английских глаголов (Rumelhart and McClelland 1986). Символисты указали на серьезные недостатки этой модели и предложили свою собственную, лучше объясняющую некоторые факты языка (Pinker and Prince 1988; Pinker 1991; Pinker and Prince 1991; Marcus et al. 1992; Prasada and Pinker 1993). В ответ на их критику коннекционисты создали пять новых вариантов сетей, штудировавших формы Прош. вр. (Plunkett and Marchman 1991, 1993; Daugherty and Seidenberg 1992; Marchman 1993; Hare and Elman 1995) и привели в своих статьях экспериментальные данные, противоречившие символистской теории, но

объяснявшиеся в рамках сетевой (Stemberger, and MacWhinney 1988; MacWhinney and Leinbach 1991; Hare, Elman, and Daugherty 1995; Elman et al. 1996; Ellis and Schmidt 1998). Это повлекло за собой серию новых публикаций со стороны символистов – они оттачивали свою модель и пытались доказать не только несостоятельность новых коннекционистских симуляций, но и ложность всего этого направления (Pinker and Prince 1994; Marcus 1995; Pinker 1995; Marcus 1998; Pinker 1999).

Почти с самого начала важную роль в этом споре играли факты, связанные с усвоением языка. Смоделировать относительно статичную систему, сложившуюся у взрослых, не в пример легче, чем постепенное развитие речи у ребенка. Основным камнем преткновения для обеих моделей оказалась так называемая "U-образная кривая" (U-shaped curve). Подражая участникам Диспута о прошедшем времени, проиллюстрируем этот феномен традиционным примером. До определенного времени англоговорящие дети успешно запоминают и употребляют формы Прош. вр. и от правильных, и от неправильных глаголов. Однако примерно в двухлетнем возрасте они вдруг начинают образовывать от неправильных глаголов формы с суффиксом *-ed* (такие, как **brea~~k~~ed*, **eate~~d~~* вместо *broke*, *ate*). Причем явление это распространяется не только на новые глаголы, но и на те, которые до этого казались совершенно освоенными. Ошибки такого рода принято называть сверхгенерализациями (overgeneralizations) – чрезмерным использованием какого-то правила или, в коннекционистской терминологии, модели. Примерно к пяти годам они исчезают.



Уже сеть Руммельхарта и МакКлеланда делала ошибки, подобные тем, которые наблюдаются в детской речи. В последующих коннекционистских моделях удалось повторить и распределение этих ошибок в виде U-образной кривой. Сделано это было следующим образом. Сначала сеть получала приблизительно равное количество правильных и неправильных глаголов. Затем количество новых правильных резко

увеличивалось. Краеугольными камнями сетевых моделей являются аналогия, частотность и обратная связь. Частотность укрепляет связи между узлами сети, по аналогии создаются новые связи. Чем частотнее какой-то класс слов, тем больше его влияние на другие. От новых и даже от уже выученных слов начинают образовываться формы по аналогии с этим классом. Это нисходящая часть кривой. Дальнейшее обучение возможно благодаря постоянной обратной связи – положительным или отрицательным реакциям "учителей" на все порожденные сетью формы. После отрицательных реакций на сверхгенерализации соотношение сил связей в сети корректируется и неправильные формы исчезают. Это восходящая часть кривой.

Возражения символистов таковы. Главное из них направлено не против конкретной модели, а против коннекционизма в целом. Это главный козырь нативизма и генеративной лингвистики – так называемый Аргумент о Бедности Стимула. Еще Платон настаивал на том, что человек не мог бы получить все знания об окружающем мире из опыта – значительная часть знаний является врожденной. Продолжая эту традицию, лингвисты-хомскианцы пытаются доказать, что в речи, которую слышит ребенок, недостаточно информации для того, чтобы построить на ее основе языковую систему⁴. Единственное объяснение того, что язык все же усваивается, – существование врожденных механизмов и знаний о том, как устроены языковые системы, то есть Универсальной Грамматики. Это очень сильный, но пока во многом спорный аргумент. Однако одну его часть генеративистам все же удалось достаточно убедительно доказать на материале лонгитюдных исследований. Большая часть родителей не исправляет грамматические ошибки в речи своих детей или делает это крайне несистематически. Более того, даже если взрослый поправляет ребенка, тот обычно не понимает, чего от него хотят. Причина заключается в том, что на первых порах ребенку очень трудно облечь в слова свои мысли и донести их до взрослых – он еще недостаточно овладел языком. Большинство взрослых

⁴ Приведем лишь один из перечисленных генеративистами доводов. Детей не учат родному языку эксплицитно, так, как взрослых учат иностранному. Все закономерности родного языка они выводят сами, причем за очень короткий срок и на таком материале, на котором их не смог бы вывести ни один взрослый человек, уже владеющий одним языком, и ни один современный компьютер. При этом в выполнении любого другого задания дети значительно уступают взрослым.

понимает это и старается не мешать и не придирается к ребенку. Исправляются только смысловые ошибки, на грамматические обычно никто не обращает внимания. Этот факт бьет в самое сердце коннекционистской модели. Ведь она исправляет все свои ошибки именно за счет того, что ей на них эксплицитно указывают. Коннекционисты пытаются показать, что взрослые, хотя и не исправляют ребенка прямо, как-то указывают ему на то, что тот сделал грамматическую ошибку. Есть и другие способы как-то обойти этот аргумент, однако все они малоубедительны.

Из менее значительных возражений против коннекционистских моделей упомянем такое. U-образная кривая в их развитии достигается за счет того, что процент правильных глаголов-стимулов вдруг резко возрастает. В естественной речи этого не происходит. Если тренировать сети в другом режиме или несколько более приблизить к реальному соотношению частотностей правильных и неправильных глаголов, они уже не смогут освоить глагольную морфологию. Есть у коннекционистских моделей и другие недостатки.

В схеме, предложенной символистами, U-образная кривая тоже оказалась слабым звеном. Предположим, что ребенок начинает с того, что запоминает формы Прош. вр. от всех глаголов, и правильных, и неправильных. При этом генетически заданная программа в его мозгу ищет в языковом материале стандартное правило. К двум годам она его находит. С этих пор стандартное правило применяется ко всем новым глаголам, если в какой-то момент не выясняется, что у них есть другая форма Прош. вр. Но как может случиться так, что от уже освоенных, то есть занесенных в лексикон неправильных глаголов также также могут образовываться формы при помощи стандартного правила? Если они уже есть в лексиконе, должно сработать блокирующее правило.

Сначала Пинкер предположил, что блокирующее правило "включается" достаточно поздно. Можно провести грубую аналогию с коренными зубами: их появление генетически задано, но вырастают они не сразу. Однако не только коннекционисты, но и единомышленники вскоре указали Пинкеру на то, что период сверхгенерализаций длится для разных глаголов разное время и даже в этот период ошибки чередуются с правильно образованными нерегулярными формами. Блокирующее правило при этом для всех одно и не может действовать выборочно. В книге, написанной Пинкером и его коллегами (Marcus et al. 1992), выдвигается гипотеза о том, что сверхгенерализации – это всего лишь ошибки

памяти. То есть, как только стандартное правило найдено, начинает работать и блокирующее. Однако при этом не всегда удается найти в лексиконе нужную нерегулярную форму. Связи между формами в детском мозгу еще не так сильны, и есть вероятность "ошибки при поиске" (retrieval failure). Сравним это с более очевидной ситуацией. Взрослый, который видел и писал букву "д" миллионы раз, вряд ли забудет, как она выглядит (хотя в шоковых ситуациях это возможно). У ребенка, который уже научился писать и читать, но еще не вполне овладел этими навыками, такое случается сплошь и рядом. В вышеупомянутой книге символисты пытаются доказать, что сверхгенерализации – это маргинальные явления. Однако их доводы не очень убедительны, так как у некоторых детей количество таких ошибок достигает 10% от всех порожденных форм и держится на таком уровне очень долго.

В целом критика односистемного подхода принципиально отличается от критики двусистемного. На данный момент все коннекционистские модели очень уязвимы. Но при этом совершенно неясно, происходит ли это из-за ложности всего направления или из-за несовершенства этих моделей. В мозгу около 100 миллиардов связанных между собой нейронов. Человек даже отдаленно не может представить себе возможности такой системы. Коннекционистские сети – очень грубое ее упрощение. Возможно, по мере того, как они будут совершенствоваться, все проблемы решатся сами собой. Модель, предложенная символистами, совсем иного толка. Ее можно опровергнуть экспериментальными данными. Ниже мы укажем несколько главных направлений, по которым сейчас ведутся эксперименты. Основная часть диплома посвящена той небольшой лепте, которую внесли в них мы.

Мы не являемся убежденными сторонниками ни одной из этих теорий – у обеих есть сильные стороны, а люди еще слишком мало знают о мозге. Так как наши предпочтения находятся на уровне интуиции, мы решили их здесь не высказывать. В любом случае, мы не думаем, что в этом споре победит какая-то из моделей, существующих на сегодняшний день, и совершенно уверены, что и та, и другая доктрина подвергнется в будущем радикальному пересмотру. Однако в силу того, что коннекционизм пока нельзя опровергнуть экспериментально, все убедительные свидетельства в его пользу – это опровержение двусистемной модели. Поэтому все: и символисты, и сторонники сетевой теории – работают именно с двусистемной моделью и

изучают, может ли она объяснить те или иные данные. По той же причине выбрали ее и мы.

3. Основные проблемы модели "Слова и Правило".

Во-первых, достоверность двусистемной модели неоднократно пытались проверить нейролингвисты. Мы коснемся этих исследований лишь вскользь, так как они меньше других связаны с тематикой нашей работы. Различия между правильными и неправильными глаголами и существительными были найдены при помощи таких техник, как позитронно-эмиссионная томография (PET), функциональное магнитно-резонансное картирование (fMRI), изучение вызванных потенциалов (ERP) (например, Jaeger et al. 1996; Ullman et al. 1997a; Clahsen 1999). Было показано, что различные болезни мозга, такие, как аграмматизм, специфическое языковое нарушение (SLI), болезнь Альцгеймера, синдром Вильямса и различные афазии также по-разному сказываются на правильных и на неправильных глаголах (Grodzinsky 1990; Van der Lely and Ullman 1996; Marslen-Wilson and Tyler 1997; Clahsen & Almazán 1998; Ullman and Gopnik 1999).

Коннекционисты поставили под сомнение результаты некоторых из этих экспериментов (например, Seidenberg and Hoeffner 1998). Однако наиболее полная их критика содержится в журнале Behavioral and Brain Sciences за 1999 год. Этот журнал в каждом выпуске публикует несколько больших статей, авторы которых должны не только представить свои новые достижения, но и показать всю панораму событий, развивающихся в этой области. Затем наиболее известным специалистам по этому вопросу, преимущественно несогласным с автором, предлагается написать небольшие критические статьи. Автор имеет возможность кратко ответить на критику. Получившаяся дискуссия публикуется на страницах журнала. Диспут о прошедшем времени был описан убежденным символистом Гаральдом Класеном (Clahsen 1999). С критикой выступило более 60 ученых, в том числе нейролингвистов. Так как Класен упомянул все перечисленные выше достижения символизма, оппоненты также не обошли вниманием ни одно из них. Кроме того, как справедливо отмечают многие коннекционисты, различия, найденные символистами, обычно могут быть объяснены в рамках односистемного подхода (например, разной частотностью регулярных и нерегулярных классов). А вот

отсутствие таких различий, продемонстрированное на экспериментальном материале некоторыми оппонентами Класена, в рамках двусистемного подхода объяснить нельзя.

Другая слабая сторона двусистемного подхода – это применимость модели Пинкера к языкам, непохожим на английский. Наверное, почти все участники Диспута о прошедшем времени хотя бы раз высказались о том, что их занятие несколько абсурдно. Споря о морфологии в ментальном лексиконе, они проверяют свои модели на морфологически нищем языке, где все глаголы разделились на 96% правильных и 4% неправильных, которые даже не образуют единого класса⁵. Эту ошибку, хотя и не сразу, попытались исправить. Например, символисты нашли стандартное правило в немецком, арабском, венгерском языках – это неполный список (McCarthy and Prince 1990; Clahsen 1999; Lukács and Pléh 1999). Коннекционисты провели эксперименты, подтверждающие его отсутствие, на материале итальянского, норвежского, исландского и других языков (Orsolini and Marslen-Wilson 1997; Orsolini Fanari and Bowles 1998; Ragnarsdóttir, Simonsen and Plunkett 1999).

Несмотря на видимые успехи обеих сторон, заметно, что коннекционисты не унывают, а ряды символистов пошатнулись. Специфика коннекционизма на сегодняшний день такова, что у его сторонников всегда есть в запасе контраргумент: то, что представляется нам в виде действия стандартного правила, может быть смоделировано и в рамках ассоциативной сети. Ни опровергнуть, ни проверить это утверждение пока невозможно. А вот для двусистемного подхода языки без единственного стандартного правила представляют большую проблему. Класен категорически настаивал на существовании такого правила в своей статье в *Behavioral and Brain Sciences*. Однако, отвечая затем на яростную критику, он должен признаться: "Важно отметить, что ничто в двусистемной модели не мешает нам предположить существование более чем одного правила на словоизменительную категорию". Действительно, основополагающая идея двусистемного подхода – это противопоставление лексического и грамматического, то есть комбинаторного компонентов. Одно или два правила в комбинаторном компоненте – это уже не так важно. Однако заметим, что модель "Слова и Правило" пока что никто не переформулировал, а ведь сам механизм ее действия строится именно на том, что правило

⁵ Так как все считают глаголы по-разному, сошлемся на источник этих цифр – Ragnarsdóttir, Simonsen and Plunkett 1999.

– одно. Класен говорит далее о возможности существования "малых" правил в таких языках, как немецкий, итальянский, польский и русский. Мы надеемся показать в этой работе, что для последнего из них это, вероятнее всего, верно (если мы находимся внутри символистского подхода и в принципе верим в правила), а следовательно, двусистемная модель срочно нуждается в пересмотре.

Есть еще одно направление, на котором символисты могут не удержаться на первоначальных категорических позициях. Одна из особенностей двусистемной модели, которая вызвала возражение многих лингвистов, – утверждение о том, что те формы, которые образуются по правилу, никогда не хранятся в памяти целиком. Из этого можно сделать вывод, что и доступ ко всем правильным формам в ментальном лексиконе (то есть их восприятие) возможен только через морфологический анализ, так как в собранном виде они в мозгу не существуют. Здесь мы попадаем в отдельную область исследований, связанных с ментальным лексиконом, – в сферу изучения лексического доступа.

На данный момент в этой области существует две основных школы. Большинство ученых считает, что лексический доступ к сложным формам может осуществляться и через их цельные образы, и посредством морфологического анализа (Taft 1979; Caramazza, Laudanna and Romani 1988; Baayen, Dijkstra and Schreuder 1997; Marslen-Wilson 2001 и другие). Из этих двух путей мозг выбирает более экономный. Некоторые лингвисты полагают, что лексический доступ всегда осуществляется через цельные образы (например, Lukatela et al. 1980). Однако лишь немногие сомневаются в том, что наиболее частотные морфологически сложные формы, в частности, формы от правильных глаголов, хранятся в памяти. Был проведен ряд экспериментов, подтвердивших эту гипотезу (например, Schreuder and Baayen 1995, Alegre and Gordon 1999).

Несмотря на то что данные большинства этих исследований были известны Класену, когда он писал статью для *Behavioral and Brain Sciences*, он все же настаивал на том, что правильные формы никогда не хранятся в памяти. Символисты провели ряд своих экспериментов, в которых пытались это доказать. Однако многие из этих экспериментов сомнительны, а ответы Класена на многочисленные возражения (Байби, Лауданны, Шрейдера и Байена и многих других) звучат довольно неубедительно. Нам кажется, что в конце концов символистам придется отказаться от этого утверждения. Это не уничтожит их теорию, но, как и в случае с несколькими правилами вместо одного,

потребуется ее пересмотра. В нашей работе приводятся материалы эксперимента на русском языке, которые также указывают на то, что частотные морфологически сложные формы хранятся в памяти целиком.

II. Проблема существования в ментальном лексиконе стандартных правил.

1. Эксперимент со взрослыми носителями: дизайн.

Глагольные классы русского языка исключительно разнообразны. Они различаются между собой с точки зрения размеров, продуктивности и сложности парадигмы⁶. Так как эти различия в основном носят градуальный характер, четкое противопоставление "правильных" и "неправильных" глаголов для русского языка нерелевантно. Целью первого эксперимента было проверить, есть ли, несмотря на все эти факторы, в русской глагольной морфологии стандартное правило, подобное тому, которое постулирует Пинкер в своей модели.

Эксперимент проходил следующим образом. Был составлен список из 48 квазиглаголов в форме Мн. ч. Прош. вр. Квазиглаголы были получены следующим образом: в основах реальных глаголов заменялось несколько звуков, несущественных с точки зрения возможных чередований, например, *плавали – клавали, играли – кидрали*. Испытуемым предлагалось устно, в форме ответов на вопросы, образовать от них две формы Наст. вр.: 3 л. Мн.ч. и 1 л. Ед.ч. Например, говорилось:

- *Вчера они клавали. А что они делают сегодня? Сегодня они...*

- *Сегодня они клаваят,* – говорил испытуемый.

- *А Вы?*

- *Сегодня я клаваяю,* – отвечал он.

Участники также имели возможность заглядывать в распечатки этих квазидialogов, где на месте их ответов были оставлены пропуски. Это было сделано для того, чтобы, с одной стороны, испытуемые не задумывались над ответами (устная форма эксперимента задавала очень быстрый темп), а с другой – чтобы им не приходилось воспринимать квазиглаголы только на слух, что очень сложно. Все материалы записывались на аудиокассеты, а потом расшифровывались и заносились в компьютер для

⁶ Под сложностью парадигмы глагола / класса глаголов мы понимаем количество и характер тех формальных преобразований, при помощи которых основа инфинитива может быть поставлена в соответствие основе Наст. вр.

дальнейшей обработки. В эксперименте приняло участие 27 русскоговорящих испытуемых (не филологов).

Прежде чем перечислить вошедшие в эксперимент глагольные классы, надо заметить, что в данной работе названия классов даются по Одноосновной Системе Р.О. Якобсона. Она представляется нам более удобной в использовании, чем традиционная система, изложенная в Академической Грамматике. В частности, в Одноосновной Системе классы не делятся на многочисленные подклассы, а их названия более наглядны. Так как в русскоязычных работах Одноосновная Система почти не используется, нам показалось целесообразным поместить в конце работы два приложения, в которых кратко излагаются основные принципы Одноосновной Системы, перечисляются и иллюстрируются примерами глагольные классы (Приложение 1), а также дается таблица соответствий между ними и теми глагольными классами, которые даны в Академической Грамматике (Приложение 2). Тем самым мы избежим необходимости постоянно разъяснять, что представляет собой тот или иной класс.

Чтобы предлагаемые испытуемым тесты имели разумную длину, в них были включены не все глагольные классы русского языка. В тесты вошли глаголы следующих классов⁷:

АЙ – 6

А – 6

ЕЙ – 6

Е – 7

(И)Й – 2

И – 8

ОВА – 6

АВАЙ – 5

ОЙ – 2

⁷ Так как квазиглаголы были получены путем незначительной трансформации реальных глаголов, за каждым из них условно сохраняется тот класс, к которому принадлежал аналогичный реальный глагол.

В таблице ниже указано количество глаголов в каждом классе и является ли он продуктивным:

	Количество глаголов	Количество бесприставочных глаголов	Продуктивность
АЙ	11735		продуктивный
А	842	91	
ЕЙ	638	252	продуктивный
Е	326	45	
(И)Й	163	5	
И	7174		продуктивный
ОВА	2815		продуктивный
АВАЙ	93	4	
ОЙ	98	5	

Эти числа взяты из созданной нами базы данных по русскому глаголу, о которой рассказывается в четвертой главе.

Первые 6 выбранных классов составляют друг с другом пары:

Классы АЙ и А – глаголы оканчиваются в прош. вр. на *-али*

Классы ЕЙ и Е – глаголы оканчиваются в прош. вр. на *-ели*

Классы (И)Й и И – глаголы оканчиваются в прош. вр. на *-или*

Определить, к какому из двух классов принадлежит каждый конкретный квазиглагол, по форме прошедшего времени невозможно. Таким образом, у испытуемых было две возможности, не противоречащих системе языка. При обсчете экспериментальных данных предстояло выявить соотношение между этими "парными" классами в ответах. Заметим, что с точки зрения величины и продуктивности эти классы делятся на огромные продуктивные классы АЙ и И, довольно небольшой, но продуктивный класс ЕЙ, небольшие непродуктивные классы А и Е и крошечный класс (И)Й.

Следующие классы, ОВА и АВАЙ, имеют специфическое суффиксальное чередование в основах Наст. и Прош. вр. (*ова//уј, ева//юј* и *ва//ј*). В класс ОВА входят глаголы с инфинитивами на *-овать* и *-евать*, однако в тесте были представлены только первые⁸. Интересно было проследить, идентифицируют ли испытуемые суффиксы в

⁸ Дело в том, что ситуация с глаголами на *-евать* значительно усложняется тем, что в русском языке есть большая группа глаголов на *-евать*, принадлежащих к классу АЙ. Кроме

формах Прош. вр. на *-овали* и *-авали* и воспользуются ли ими как морфологической подсказкой при определении глагольного класса. Если суффиксы останутся незамеченными, от этих глаголов можно будет образовать формы Наст. вр. по модели АЙ-класса, подобно глаголу *читать – читают, читаю*. Заметим, что суффиксы классов ОВА и АВАЙ хорошо опознаются на слух. Более того, из всего множества глаголов на *-авать* только один единственный глагол *упавать* относится не к классу ОВА, а к классу АЙ. Глаголов на *-авать* в русском языке немного, зато все они принадлежат к классу АВАЙ. Это глагол *давать* и его производные, глагол *создавать*, а также различные префиксальные глаголы на *-ставать* и *-знавать*.

Класс ОЙ был выбран потому, что форма Мн. ч. Прош. вр. у глаголов данного класса оканчиваются на *-ыли*, что в русском языке встречается крайне редко. Это характерно для В-класса, к которому относятся глаголы *жить, плыть, слыть* и их производные, собственно для ОЙ-класса, включающего 5 непрефиксальных глаголов (*выть, крыть, мыть, ныть, рыть*) и их производные, а также для глагола *стыть*, который в Академической Грамматике рассматривается как вариантная форма по отношению к *стынуть*, а по Одноосновной Системе включается в Н-класс. Класс В почти не принимался нами в расчет, так как соотношение основ Наст. вр. и инфинитива, которое в нем наблюдается⁹, совершенно нехарактерно для системы русского языка в целом, а сам класс ничтожно мал. Класс ОЙ, тоже непродуктивный и крайне немногочисленный, все же лучше вписывается в систему глагольных классов и содержит больше глаголов. Основа Наст. вр. у глаголов этого класса образуется путем прибавления к основе инфинитива *|j|* и чередования корневых гласных *o//ы*. Эксперимент должен был показать, как поведут себя испытуемые в той ситуации, когда использование "внутрисистемных" возможностей предполагает применение достаточно сложного и редкого алгоритма.

Специфика этого эксперимента, прежде всего, заключается в том, что перед нами не стоял вопрос, какие формы испытуемые извлекли из памяти, а какие – сконструировали. Задание с квазиглаголами имитирует ситуацию столкновения с новым

глаголов *издеваться, намереваться, недоумевать, обуревать, подозревать, сомневаться, увещевать*, это также множество глаголов, в которых *-ва-* является суффиксом имперфективации, например, *запевать* или *загустевать*.

⁹ *плы-л-и — плыв-ут*

незнакомым глаголом. Это значит, что нужная форма не может быть извлечена из памяти. В символистской модели новые формы могут либо образовываться при помощи стандартного правила, как формы слов регулярного класса, либо создаваться по аналогии с нерегулярными формами. Символисты часто используют эксперименты с квазиглаголами, так как в них наиболее наглядно может проявиться разница между предполагаемым правилом и аналогическими процессами (несколько экспериментов разобрано в Clahsen 1999 Класеном и его оппонентами). Отличие правила заключается в том, что оно может быть сверхгенерализовано, то есть применено практически к любому слову. По аналогии могут быть образованы только немногочисленные формы от тех квазислов, которые фонетически сходны с прототипами¹⁰.

Заметим, что в условиях устного эксперимента с наивными носителями языка "конструирование" форм от квазиглаголов, видимо, происходит почти или совершенно бессознательно. Вопрос о том, какие именно процессы идут при этом у человека в мозгу, мы затрагивать ни в коем случае не будем. Говоря "испытуемый выделяет основу прошедшего времени" или "испытуемый выбирает глагольный класс", мы не имеем в виду ни сознательные действия человека, ни последовательности операций, совершающихся у него в мозгу. Мы просто констатируем результаты определенных преобразований, произошедших с исходной формой. Совокупности этих преобразований мы будем называть моделями, под которыми будем понимать использованные испытуемыми различные способы выделения основы Прош. вр. и превращения ее в основу Наст. вр.

Классы	Примеры	Модели (формальное описание соответствия между пр. вр. и наст. вр.)
АЙ	<i>делали – дела[jy]т, дела[jy]</i>	Прош. вр. – LI + J + UT / JU
ЕЙ	<i>имели – име[jy]т, име[jy]</i>	Прош. вр. – LI + J + UT / JU
(И)Й	<i>гнили – гни[jy]т, гни[jy]</i> <i>пили – п[jy]т, п[jy]</i>	Прош. вр. – LI + J + UT / JU Прош. вр. – LI + (I)J + UT / JU
А	<i>писали – пишут, пишу</i>	Прош. вр. – А – LI + UT / U, черед. согл. в 1 л. Ед. ч. и 3 л. Мн. ч.
Е	<i>видели – видят, вижу</i>	Прош. вр. – Е – LI + AT / U, черед. согл. в 1 л. Ед. ч.
И	<i>просили – просят, прошу</i>	Прош. вр. – I – LI + AT / U, черед. согл. в 1 л. Ед. ч.
ОВА	<i>требовали – требу[jy]т, требу[jy]</i>	Прош. вр. – LI + J + UT / U ОВА // U

¹⁰ То есть возможность использования аналогии зависит от фонологического окружения.

АВАЙ	<i>давали – да[jy]т, да[jy]</i>	Прош. вр. – LI + J + UT / U VA // J
ОЙ	<i>мыли – мо[jy]т, мо[jy]</i>	Прош. вр. – LI + J + UT / U Y // O

Заметим, что, анализируя разные глагольные классы на предмет сложности парадигмы, следует учитывать, что важно не только количество правил, задействованных в парадигме, но и то, что это за правила. Известно, например, что некоторые чередования согласных иногда представляют трудности даже для взрослых носителей русского языка, образующих формы от реальных глаголов. Таким образом, если в парадигму входит такое правило, сложность ее должна увеличиваться.

2. Эксперимент со взрослыми носителями: результаты.

В тесте имелись квазиглаголы, полученные из реальных глаголов 9 различных классов. Образовывая от них формы Наст. вр., испытуемые применяли определенные модели, свойственные тем или иным глагольным классам языка. В сущности, это были те же самые 9 классов, только их соотношение в ответах было совершенно иным, чем в списке стимулов: во многих случаях квазиглаголы, соотносящиеся с реальными глаголами одного класса, были интерпретированы как глаголы другого класса. В ответах испытуемых не встретилось не вошедших в эксперимент глагольных классов.¹¹ Результаты эксперимента приводятся ниже в форме таблиц:

		РЕАКЦИИ										Другие ответы	Всего
		АЙ	А	(И)Й	И	Е	ЕЙ	ОВА	АВАЙ	ОЙ	(ЫЙ)		
С Т И М У Л Ы	АЙ	287	2		11		10					10	320
	А	262	38				10					14	324
	(И)Й			26	18		3	9				52	108
	И	11		127	156		52	1			2	77	426
	Е	5				96	227	3				45	376
	ЕЙ	2		2		24	239					57	324

¹¹ В этом факте нет ничего удивительного, если принять во внимание, что те классы, которые не были включены в эксперимент, имеют характерные формы Прош. вр., отличные от тех, которые были представлены в тесте.

ОВА	131			5			153				35	324
АВАЙ	165			1		8	25	59			12	270
ОЙ	1			7		1			1	51	47	108

		РЕАКЦИИ					
		АЙ	А	(И)Й	И	Е	ЕЙ
СТИМУЛЫ	АЙ	89.7%	0.6%		3.5%		3.1%
	А	80.9%	11.7%				3.1%
	(И)Й			24.1%	16.7%		2.8%
	И	2.6%		30.0%	36.6%		12.2%
	Е	1.3%				25.5%	60.4%
	ЕЙ	0.6%		0.6%		7.4%	73.8%
	ОВА	40.4%			1.6%		
	АВАЙ	61.1%			0.4%		3.0%
	ОЙ	0.9%			6.5%		0.9%

		РЕАКЦИИ				
		ОВА	АВАЙ	ОЙ	(ЫЙ)	Другие ответы
СТИМУЛЫ	АЙ					3.1%
	А					4.3%
	(И)Й	8.3%				48.1%
	И	0.2%			0.4%	18.0%
	Е	0.8%				12.0%
	ЕЙ					17.6%
	ОВА	47.2%				10.8%
	АВАЙ	9.3%	21.8%			4.4%
	ОЙ			0.9%	47.3%	43.5%

	РЕАКЦИИ											Всего
	АЙ	А	(И)Й	И	Е	ЕЙ	ОВА	АВАЙ	ОЙ	(ЫЙ)	Другие ответы	
на -али	549	40		11		20					24	644
на -или	11		153	174		55	10			2	129	534
на -ели	7		2		120	466	3				102	700

	РЕАКЦИИ										
	АЙ	А	(И)Й	И	Е	ЕЙ	ОВА	АВАЙ	ОЙ	(ЫЙ)	Другие ответы
на -али	85.2%	6.2%		1.7%		3.1%					3.8%
на -или	2.1%		28.6%	32.6%		10.3%	1.9%			0.4%	24.1%
на -ели	1.0%		0.3%		17.1%	66.6%	0.4%				14.6%

Таблицы требуют некоторых пояснений. В некоторых случаях модели, использованные участниками эксперимента, вообще не имели аналогов в языке. Большая часть таких ответов трудно поддается систематизации и не дает нам никакой существенной информации. Это такие формы, как, например, *колядуют* (стимул *кледава́ли*, от реал. гл. *продава́ли*) или *тра́калеют* (стимул *тра́кали* от реал. гл. *пла́кали*). Эти ответы заносились нами в графу "Другие ответы". Однако некоторые формы, образованные не по правилам языковой системы, представляли для нас большой интерес и были рассмотрены особо. Это формы типа *брЫют* (стимул *брЫли*, от реал. гл. *крыли*). Число ответов с такой моделью попало в отдельную графу, названную (ЫЙ). Заметим также, что для некоторых испытуемых известны ответы не на все 48 вопросов теста (несколько раз один глагол бывал пропущен, в двух местах есть дефекты записи). Это отразилось на общем числе ответов на тот или иной стимул.

На этапе общей статистической обработки была выявлена модель, которая используется испытуемыми существенно чаще, чем все остальные. Вот ее формальное описание:

$$\text{Наст. вр.} = \text{Прош. вр.} - \text{ЛI} + \text{J} + \text{UT} / \text{JU}$$

Назовем ее для краткости Й-моделью. В языке эта модель используется в продуктивных классах АЙ и ЕЙ и в крошечном (И)Й-классе. Заметим, что сложность парадигмы у глаголов с такой моделью минимальная.

Однако в русском языке есть и другие продуктивные глагольные классы с высокой частотностью, в которых используются другие модели. Наличие стандартного правила в системе глагольных классов определяется не существованием модели с наибольшей частотностью и наименьшей сложностью парадигмы. Можем ли мы называть ту или иную модель стандартной, определяется ее независимостью от фонетического окружения и мерой ее влияния на всю систему, на все глагольные классы. Чтобы показать, что в

ответах испытуемых такое влияние действительно наблюдается и что оно довольно сильно, мы разберем результаты по всем вошедшим в тест глагольным классам.

85,2% квазилаголов на *-али* были опознаны испытуемыми на как принадлежащие к АЙ-классу, и всего лишь 6,2% – как принадлежащие к А-классу. От 66,6% квазиглаголов на *-ели* было образовано Наст. вр., как от глаголов ЕЙ-класса, а от 17,1% – как от глаголов Е-класса. В том, что классы АЙ и ЕЙ встречались в ответах чаще, не было ничего неожиданного. Однако то, что число ответов по моделям А-класса и Е-класса будет настолько мало, совсем не было очевидно. Конечно, это небольшие и непродуктивные классы, однако к ним принадлежит очень много высокочастотных глаголов.

Квазиглаголы на *-или* на 32,6% были опознаны как глаголы И-класса и на 28,6% – как глаголы (И)Й-класса¹². Этот результат можно назвать неожиданным, если учесть, что в русском языке почти все глаголы на *-или* принадлежат к продуктивному И-классу. И-класс характеризуется очень высокой частотностью. То, что в ответах испытуемых было так мало форм, образованных по модели И-класса, и так много – по Й-модели, хотя в языке глаголов на *-или* с такой моделью очень мало, может быть объяснено только влиянием некоего надсистемного правила, то есть существованием дефолта.

Перейдем к квазиглаголам на *-овали*. Как уже было сказано выше, в русском языке существует всего один глагол на *-овали*, образующий формы по Й-модели: глагол *уповать*. Все остальные многочисленные глаголы относятся к ОВА-классу. В формах Прощ. вр. на принадлежность глаголов к ОВА-классу указывает соответствующий суффикс – морфологическая подсказка, которую сложно не заметить любому носителю языка. Однако многие испытуемые восприняли квазиглаголы *-овали* просто как глаголы на *-али*, проигнорировав суффикс *-ова-*, и применили к ним Й-модель (как в АЙ-классе). Й-модель лишь незначительно уступает той модели, которая свойственна глаголам на *-овали* в языке (40,4% и 47,2%).

Квазиглаголы на *-авали*, также содержавшие морфологическую подсказку в виде суффикса, были опознаны как глаголы АВАЙ-класса всего в 21,8% случаев, хотя в языке все глаголы на *-авали* входят в АВАЙ-класс. Правда, таких глаголов в русском языке

¹² Эти глаголы вызвали бóльшие трудности, чем глаголы на *-али* и *-ели*. Было много форм Наст. вр., в которых был оставлен суффикс прошедшего, и так далее. Все они попали в графу "Другие ответы".

мало: *создавать*, *давать* и приставочные глаголы с этим корнем, а также глаголы на *-ставать* и *-знавать*. Однако многие из этих глаголов очень употребительны. Незначительная часть квазиглаголов на *-авали* (9,3%) была отнесена к ОВА-классу – это объясняется редукцией безударных гласных в русском языке. А от 61,1% квазиглаголов были образованы формы по Й-модели, чего в реальном языке вообще не встречается.

То, что испытуемые проигнорировали столь явные морфологические подсказки и применили к квазиглаголам модель, совершенно нехарактерную для подобных им реальных глаголов, также указывает на то, что в языке существует стандартное правило и правило это – Й-модель.

Теперь обратимся к квазиглаголам на *-ыли*. Реальные глаголы на *-ыли* принадлежат к двум маленьким непродуктивным классам (Н-классу и ОЙ-классу). Квазиглаголы, вошедшие в наш тест, были образованы путем трансформации глаголов ОЙ-класса. Модель Н-класса, к которому принадлежат глаголы *жить*, *плыть*, *слыть*, испытуемые не использовали ни разу. Количество форм, образованных по модели ОЙ-класса, к которому относятся глаголы *выть*, *крыть*, *мыть*, *ныть*, *рыть*, составило 0,9% ответов. 47,3% ответов было образовано по Й-модели, что с точки зрения реального языка совершенно недопустимо. Этот факт также подтверждает нашу гипотезу относительно стандартного правила.

Если отвечавший воспользовался моделью, отличной от Й-модели, было прослежено, “грамотно” ли он ее применил: появились ли чередования и правильно ли выбрано спряжение? Ведь в случае регулярных ошибок при использовании тех или иных моделей можно утверждать, что правила, которые образуют эту модель в языке, применялись испытуемыми не все вместе, как единое макроправило, а по отдельности. Чередования очень часто игнорировались тестируемыми. Спряжение же в большинстве случаев было выбрано верно. То есть тогда, когда от глаголов на *-или* и *-ели* образовывались формы как от глаголов И-класса и Е-класса, эти формы имели окончания второго спряжения.

Далее нам хотелось бы остановиться на одном интересном факте. Некоторые квазиглаголы вызвали сильные трудности почти у всех участников эксперимента. В основном это были двусложные квазислова типа *кíли* (от реал. гл. *пили*) или *зЫли* (от реал. гл. *мыли*). Многие испытуемые сначала затруднялись ответить, а в итоге оставляли в

формах Наст. вр. суффикс Прощ. вр. *-л-*. Особенно тяжело пришлось квазилаголам на *-ыли*, у которых должно быть чередование гласных в основах Прощ. и Наст. вр.

Нам кажется, что здесь косвенно сказывается существующее в языке соотношение между конструируемыми и извлекаемыми из памяти формами и напрямую связанные с ним представления о том, как может выглядеть новый незнакомый глагол. Начиная с определенного возраста подавляющее большинство новых глаголов, с которыми сталкивается носитель языка, – это "длинные" глаголы, принадлежащие к продуктивным классам. Поэтому в тесте все без особого труда справляются с трехсложными глаголами, к которым можно применить "хорошо отработанную" модель, свойственную одному из продуктивных классов. Встретить незнакомый "короткий" глагол для взрослых носителей языка почти нереально, все короткие глаголы уже давно известны. Видимо, про эти глаголы мы можем утверждать, что они наверняка в том или ином виде хранятся в памяти целиком, так как, судя по тесту, обращаться с такими новыми глаголами испытуемые просто не умеют. Вот какие ответы давали участники эксперимента на стимул *зЫли* (от реал. гл. *мыли*):

<i>зЫют, зЫю</i>	8
<i>зЫлят, зЫлю</i>	5
<i>зылѐют, зылѐю</i>	2
<i>злют, злЮ</i>	2
<i>зыЮт, зыЮ</i>	2
<i>зЯт, зЮ</i>	2
<i>знѹт, знѹ</i>	1
<i>знЫют, знЫю</i>	1
<i>зЫлют, зЫлю</i>	1
<i>зылЮют, зылЮю</i>	1
<i>зЫют, зыЮ</i>	1
<i>зЯт, зѐю</i>	1

До сих пор речь шла об ответах, более или менее "вписывающихся" в систему языка. Под этим мы подразумеваем то, что при помощи ряда операций "отнять-прибавить", примененных к квазиглаголу, в языке хотя бы теоретически может быть получена форма Наст вр. реального глагола какого-нибудь класса из формы Прощ. вр. Но среди ответов встретились и другие формы. Отсечение от исходной формы 3 л. Мн. ч. суффикса Прощ. вр. *-л-* и окончания *-и* происходило не всегда: многие испытуемые, образуя форму Наст. вр., оставляли в основе глагола этот суффикс, а иногда даже и

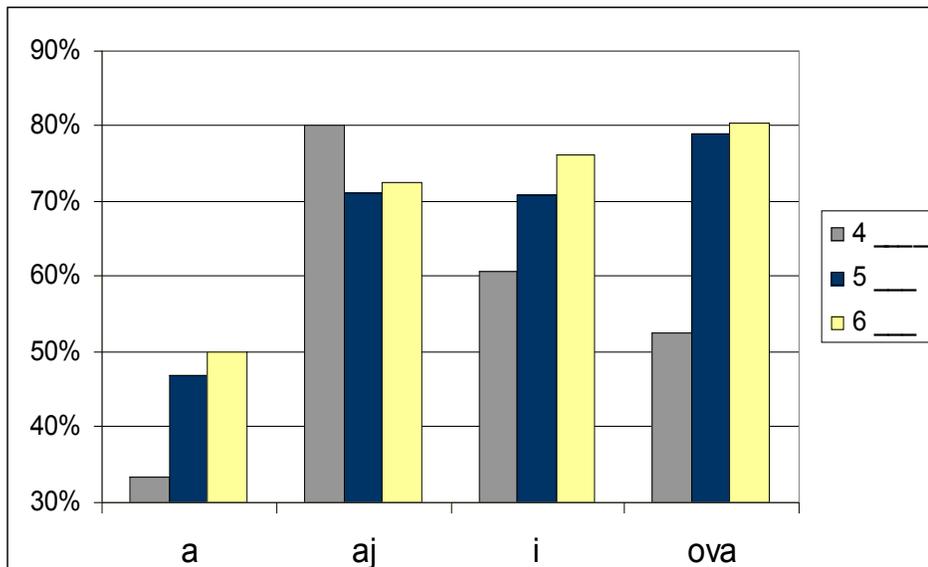
окончание. Значит, показатель Прош. вр. *-ли* совершенно не обязательно автоматически воспринимался тестируемыми как таковой, раз в ряде случаев он был расценен как неотъемлемая принадлежность глагольной основы в целом. Этот факт кажется нам крайне интересным, однако пока не совсем понятно, каким образом можно анализировать конкретные формы. Далее с полученной "псевдоосновой" производились разнообразные операции: кроме окончаний настоящего времени к ней могли быть присоединены суффиксы, иногда совершенно неожиданные, скажем, *-ва-* или *-у-* и *-j-*, так что, например, от квазиглагола *угели* была образована форма *угеливаю*. Надо заметить, что в большинстве случаев отвечавшие предпочитали такую модель, когда после *-л-*, оставшегося от прошедшего времени, следовал некий гласный, */j/* и окончания, например, от квазиглагола *бнили* – форма *бнилею*. Такие формы являются, в сущности, результатом применения нашей Й-модели к неправильно определенной основе настоящего времени.

3. Эксперименты с детьми: сравнение результатов.

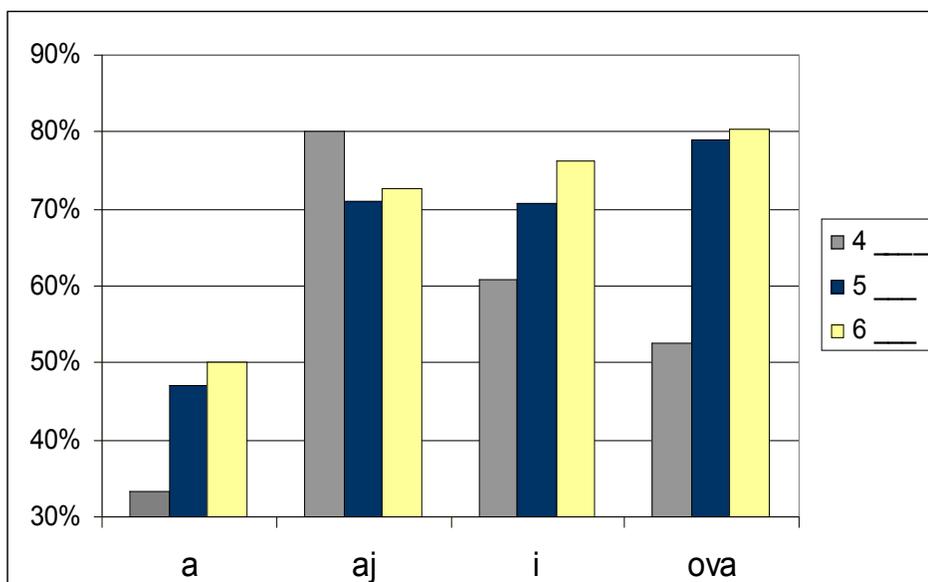
Итак, мы показали, что несмотря на разнообразие глагольных классов и на кажущееся равноправие наиболее частотных из них, в русском языке может быть найдено правило, которое ведет себя как стандартное. А ведь когда этот эксперимент только задумывался, казалось, что данные русского языка не оставят камня на камне от символистской модели. После эксперимента со взрослыми другие студенты (главным образом, Татьяна Свистунова) провели в рамках нашего проекта серию экспериментов с детьми.

Задание было то же, что и в первом эксперименте, только глагольных классов было меньше, и часть стимулов была реальными словами. Анализ данных показал, что у детей также наблюдаются сверхгенерализации по Й-модели, которая предположительно является стандартным правилом (если правила вообще существуют). Пик таких сверхгенерализаций приходится на 4 года или на более ранний возраст (в нашем проекте четырехлетки были самыми маленькими испытуемыми). Однако позже у детей появляются сверхгенерализации другого типа – по модели ОВА класса (они наиболее часто встречаются у пятилетних). Судя по всему, именно в это время дети осваивают ОВА класс (количество правильных ответов на реальные глаголы и квазиглаголы этого класса

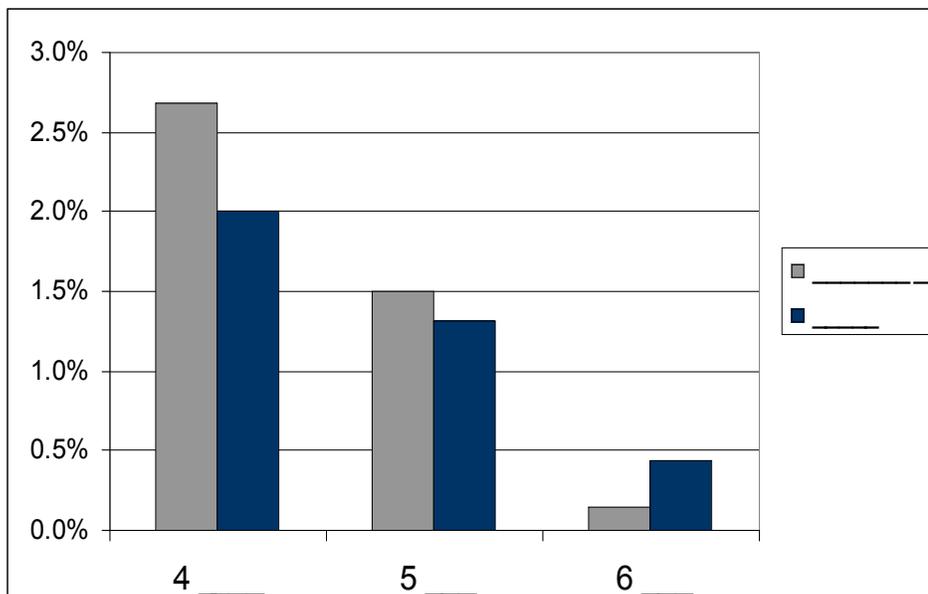
резко возрастает). Вот диаграмма, на которой представлено количество правильных ответов на все стимулы (для квазиглаголов слово "правильный" условно):



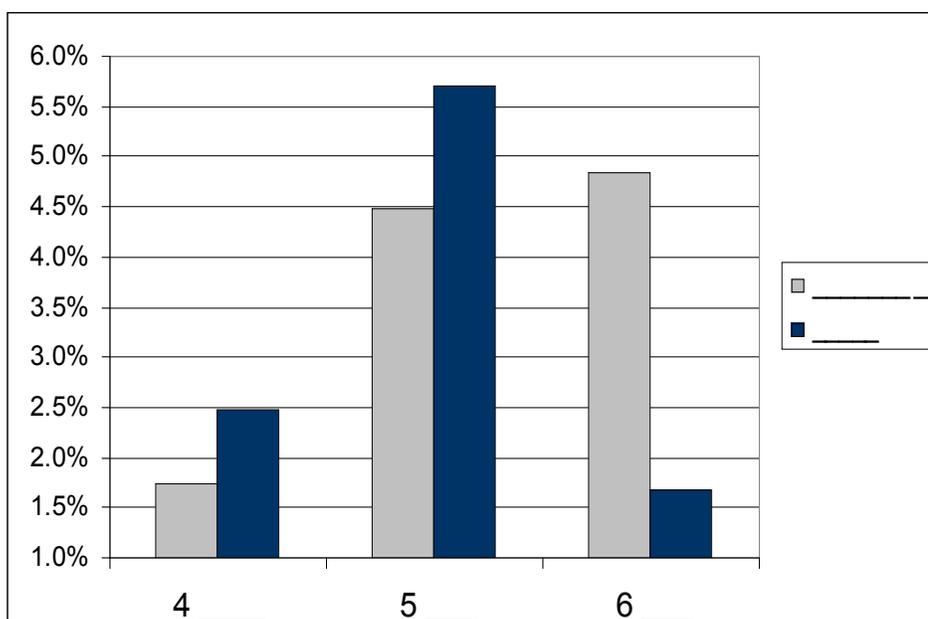
На этой диаграмме изображены проценты правильных ответов на реальные глаголы:



На следующем графике показано число форм от стимулов И-класса, образованных по стандартной модели:



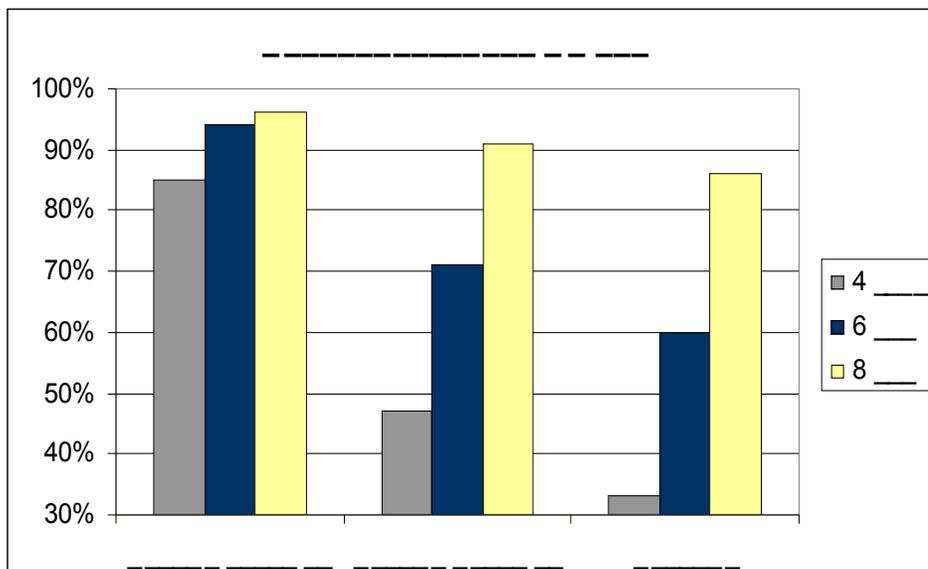
Здесь показаны свержгенерализации по модели ОВА-класса:

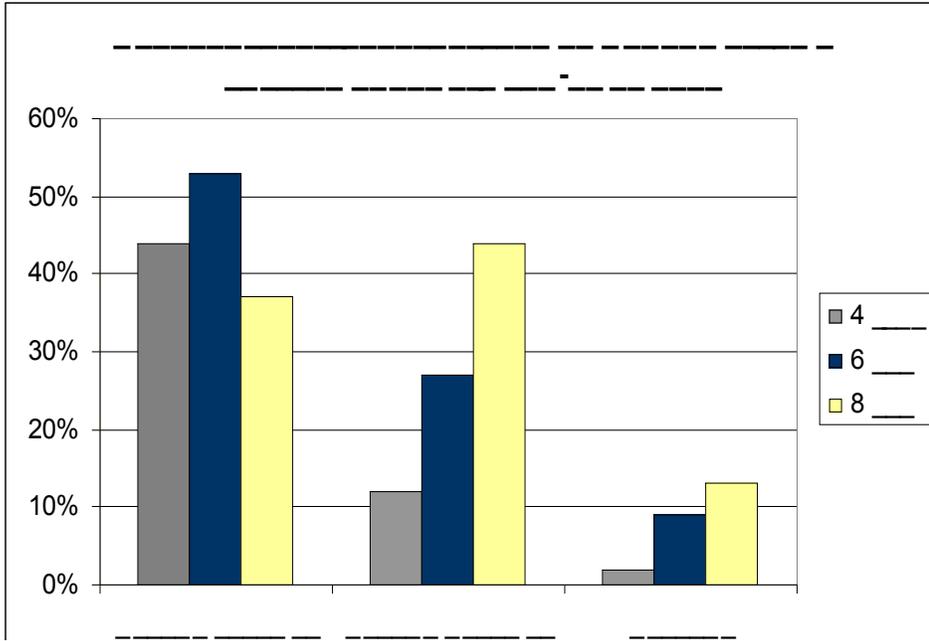


Получается, что, вопреки схеме, предложенной Пинкером, дети не запоминают глаголы ОВА-класса по одному, пользуясь аналогией, а выводят соответствующее правило. Ведь заметное количество свержгенерализаций может объясняться только этим. Для сравнения, в английском языке неправомерная аналогия с неправильными глаголами дает около 0,2% ошибок. В таком случае можно сказать, что найдено одно из тех "малых правил", о возможности существования которых упоминал Класен. Чтобы иметь

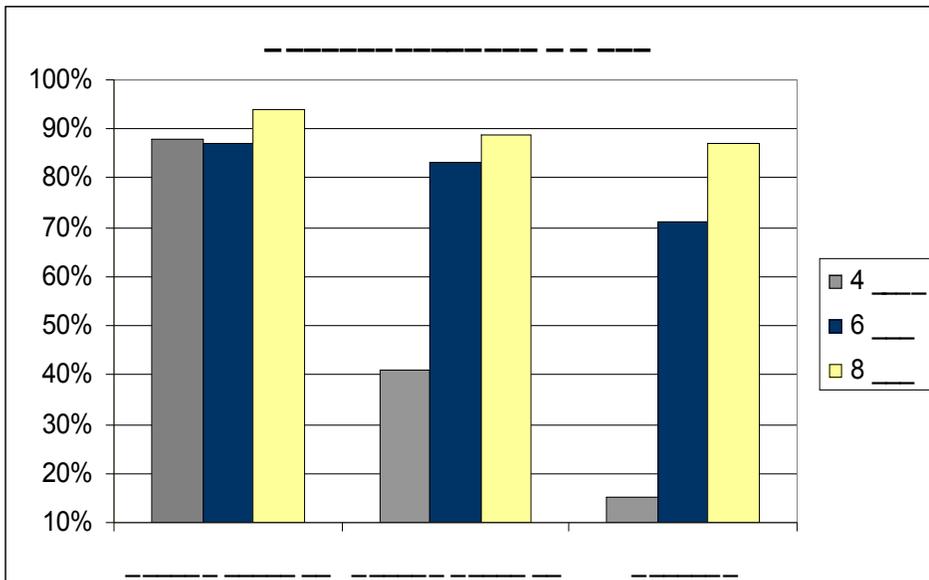
возможность объяснить эти данные, модель "Слова и Правило" должна быть сильно реорганизована и превращена в "Слова и Правила".

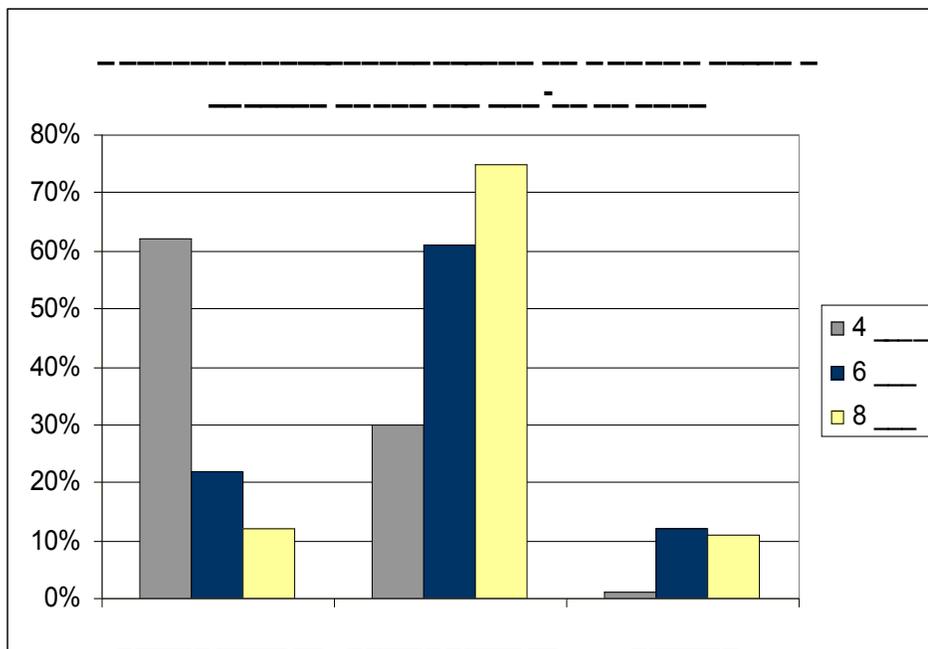
Напоследок заметим, что подобного рода "малое правило" было обнаружено и другими исследователями, изучавшими формы Прош. вр. от норвежских и исландских глаголов (Ragnarsdóttir, Simonsen and Plunkett 1999). Однако, так как они работали в рамках коннекционистской парадигмы, они не называли этот феномен правилом. В обоих языках есть неправильные, или сильные, и две группы правильных, или слабых, глаголов: большая и меньшая. Этапы овладения глагольной морфологией в обоих языках одни и те же, однако несколько смещены во времени. Сначала дети осваивают большую группу правильных глаголов (это совпадает с пиком сверхгенерализаций по этой модели), затем меньшую. В этот период появляется большое количество сверхгенерализаций по модели меньшего слабого класса. Затем осваиваются сильные глаголы. Ошибочные употребления характерных для них моделей встречаются, но довольно редко. Таким образом, в рамках символистской теории они были бы отнесены не к сверхгенерализации правила, а к неправомерному применению аналогии. Ниже помещены диаграммы с ответами норвежских детей:





А вот диаграммы, на которых показаны данные исландских детей:





В заключении авторы статьи, в числе прочего, пишут: "Неожиданным результатом оказалось то, что меньший слабый класс стал главным источником ошибок в норвежском и преобладающим источником ошибок в исландском у детей старшей возрастной группы". Так как экспериментальные данные по русским детям принесли такой же "неожиданный результат", можно говорить об очень интересной тенденции, на которую символистам стоило бы обратить внимание.

III. Проблема соотношения конструируемых и анализируемых форм и форм, хранящихся в памяти.

1. Классический эксперимент Алегре и Гордона.

Как мы уже отмечали выше, другая особенность символизма, спровоцировавшая критику многих лингвистов, заключалась в том, что в двусистемной модели те формы, которые образуются по правилу, никогда не хранятся в памяти целиком. Один из наиболее известных экспериментов, доказавших обратное, был описан в статье Марии Алегре и Питера Гордона (Alegre and Gordon 1999). Он был спланирован следующим образом. Испытуемым на экране компьютера предъявлялись слова и квазислова. После каждого стимула человек должен был ответить, слово он видел или не слово, нажав на одну из двух клавиш. Алегре и Гордон использовали в своем эксперименте морфологически сложные существительные (Мн. ч.) и глаголы (формы Прош. вр., 3 л. Ед. ч. Наст. вр., причастия), а также морфологически неделимые контрольные стимулы и различные отвлекающие слова.

Остановимся на формах Прош. вр. от правильных глаголов. Все они были объединены попарно следующим образом. Частота гнезда, то есть сумма частот всех форм, для двух этих слов была одинакова, а собственно частота формы Прош. вр. различалась. Например, сумма частот всех форм для слов *arrive* и *publish* совпадает (108 на миллион), но форма Прош. вр. *arrived* встречается реже, чем *published* (62 : 89). Если при восприятии этих форм происходит морфологический анализ, они должны распознаваться с равной скоростью, так как общая частота основ у этих лексем одинаковая. Если же эти формы хранятся в памяти целиком, реакция на форму *published* должна быть быстрее, чем на *arrived*, так как последняя менее частотна. Алегре и Гордону удалось показать, что такая разница во времени реакции существует для форм с частотностью более 6 на миллион.

Из материалов Алегре и Гордона можно сделать вывод, что формы Прош. вр. от многих английских правильных глаголов, судя по всему, хранятся в памяти целиком. Русский язык, безусловно, отличается от английского большим разнообразием форм и отсутствием единственного правильного класса. С другой стороны, мозг у носителей обоих языков устроен одинаково. Мы не знаем, существует ли для мозга некий

универсальный частотный порог, после которого выгоднее запоминать форму, а не анализировать ее. Может быть, наоборот, существует какое-то ограничение на количество запоминаемых форм (тогда носитель английского языка хранит в памяти более редкие формы, чем носитель русского). А может быть, мозг экономит главным образом усилия по конструированию или анализу форм, и тогда носители таких языков, как русский, могут хранить в памяти и более редкие формы, чем носители английского языка, так как в русском намного более сложная морфология.

2. Почему невозможно повторить эксперимент Алегре и Гордона с русскими глаголами?

К сожалению, на данный момент не представляется возможным повторить эксперимент Алегре и Гордона на материале интересующих нас русских глаголов. Во-первых, непонятно, как для русского глагола следует рассчитывать частоту гнезда. Например, существует мнение, что некоторые видовые пары представлены в ментальном лексиконе в виде некой гиперлексемы, а не как два отдельных слова. Ряд экспериментов проводила в этой области М.В. Русакова. Для некоторых глаголов, таких, например, как *срѣзать* – *срезать*, эта гипотеза с очень большой долей вероятности верна. Однако для видовых пар, устроенных иначе, это уже менее очевидно. Наконец, существуют видовые пары, по отношению к которым это утверждение явно неверно.

Далее, неясно, надо ли учитывать приставочные глаголы вместе с бесприставочным, от которого они происходят. С одной стороны, иногда приставочные глаголы несомненно являются самостоятельными словами, слабо ассоциирующимися с тем глаголом, от которого они когда-то образовались. С другой стороны, многие приставочные глаголы тесно связаны со своей бесприставочной парой. Их значение – это, по сути, сумма значения бесприставочного глагола и грамматикализованного значения, привносимого префиксом. Например, *запеть* = *петь* + значение начинания действия, закрепленное за приставкой *за-*. Такие пары, как *делать* – *сделать*, многие вообще склонны считать видовыми парами. Теоретическая русистика решает эту проблему формально: все эти глаголы считаются отдельными лексемами. Такой принцип вполне оправдан – он вносит в систему единообразие, на нем можно основываться при составлении словарей и грамматик. Однако с точки зрения психолингвистики такой

подход ничего не дает: ведь если то, что мы считаем отдельными глаголами, объединено в гиперлексему или как-то иначе тесно связано в ментальном лексиконе, возможно, для них надо считать суммарную частоту гнезда или же вводить какой-то другой, более сложный способ определения частотности.

Мы также не знаем, в каких случаях и надо ли вообще учитывать при определении частоты гнезда причастия и деепричастия. Ведь в языке они часто адъективируются или адвербиализуются, отделяясь таким образом от глагольной лексемы.

Впрочем, можно, в конце концов, остановиться на каких-то формальных критериях и самостоятельно рассчитать частотности глаголов по одному из существующих корпусов русского языка или попросту довериться составителям частотных словарей. Но это не единственная проблема. Если мы начнем подбирать пары глаголов, как Алегре и Гордон, окажется, что в русском языке намного больше характеристик, по которым они должны совпасть: класс, вид, отсутствие приставки... Чтобы включить в эксперимент глаголы разных классов, желательно, чтобы они были приблизительно одинаковой длины, но в одних классах частотные глаголы преимущественно двусложные, а в других – трехсложные. После долгих поисков мы пришли к выводу, что достаточное количество частотных глаголов для такого эксперимента набрать невозможно.

С редкими русскими глаголами невозможно работать по другой причине: для русского языка нет хороших больших частотных словарей. На сегодняшний день самые большие корпуса для русского языка насчитывают всего миллион словоупотреблений. По мнению Леннарта Лённгрена, составившего один из лучших русских частотных словарей, по таким корпусам можно делать более или менее достоверные выводы о словах (или формах) с частотностью более 80 на миллион (Lönngren, 1993). Понятно, что в подобной ситуации о таких частотностях, как 6 на миллион, не может быть и речи.

3. Подготовка нового эксперимента на материале русского языка.

Несмотря на вышеперечисленные теоретические и практические проблемы, которые не решены в современной русистике, нами были спланирован и проведен эксперимент, затрагивающий те же проблемы, что и работа Алегре и Гордона, однако несколько иного рода, чем у них. Алегре и Гордон брали формы разной частотности от глаголов с одинаковой частотой гнезда и показывали, подвергаются ли они при

восприятию морфологическому анализу. Оказалось, что это зависит от частоты гнезда: неправильные формы никогда не подвергаются морфологическому анализу, а из правильных подвергаются только редкие. Мы не можем быть уверены, что правильно рассчитываем частоту гнезда, не можем набрать достаточно пар для относительно частотных глаголов, не можем работать с редкими. Поэтому мы решили сравнить разные формы внутри одной лексемы. Таким образом снимается проблема вычисления частоты гнезда и поиска пар с одинаковой частотой. Для того, чтобы спланировать такой эксперимент, сперва пришлось провести следующее предварительное исследование.

Из частотного словаря Лённгрена были взяты все глаголы с частотностью больше 80 на миллион. В данном случае мы используем эту цифру как некий формальный критерий, поэтому нас не беспокоит тот факт, что ее психолингвистическая сущность может быть очень и очень сомнительной. Установив такой порог, мы, с одной стороны, получаем достаточно большую выборку (213 глаголов), а с другой стороны, согласно мнению самого Лённгрена, берем только такие слова, о которых его словарь дает более или менее адекватное представление.

Далее мы взяли два корпуса русского языка. Один из них – Уппсальский корпус, база частотного словаря Лённгрена, а другой – Тюбингенский корпус разговорных текстов. В Уппсальский корпус вошли отрывки художественных, научных и публицистических текстов второй половины XX века. Добротный, составленный с применением новейших статистических методов на основе относительно современных текстов, на данный момент он считается одним из лучших корпусов русского языка. В основу Тюбингенского корпуса легла подборка из интервью и дискуссий, напечатанных в различных русских газетах и журналах. В корпус вошли тексты, относящиеся к 1996 году, и более поздние (он постоянно пополняется). На настоящий момент этот корпус насчитывает более 600 000 словоупотреблений. Выбор этого корпуса объясняется следующими причинами. Многие психолингвисты, оперирующие данными о частотности тех или иных слов или форм, нередко подвергаются суровой критике. Ведь такие данные чаще всего берутся из частотных словарей, материалом для которых служат исключительно письменные, а не устные тексты. Даже неспециалисту интуитивно ясно, что именно частотность слов и их форм в разговорной речи должна коррелировать с частотностью, представленной в ментальном лексиконе. При этом мы знаем, что

частотность тех или иных слов или форм в письменной и устной речи может существенно различаться. Итак, чтобы избежать подобных упреков, мы взяли в пару к Уппсальскому корпусу Тюбингенский корпус разговорных текстов. Он, безусловно, намного менее упорядочен, и за ним не стоит такого мощного статистического аппарата. Зато он позволит нам "проверить" данные Уппсальского корпуса на разговорной речи. Оба корпуса доступны в Интернете на сайте Тюбингенского университета (<http://heckel.sfb.uni-tuebingen.de/cgi-bin/koren.pl?hello=hello&trans=lat>). Так как Тюбингенский корпус существует исключительно в электронном виде, его описание можно найти только в Интернете (<http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html#interview> и другие страницы, посвященные этому проекту).

Для выбранных 213 глаголов были вычислены частотности всех финитных форм в каждом из корпусов. Затем было подсчитано, какой процент составляют различные формы от общего количества финитных форм данного глагола. Например, возьмем глагол *выступать*. Форма Ж. р. Прош. вр. *выступала* встречается в Уппсальском корпусе 5 раз, а в Тюбингенском – 2 раза. Сумма частот всех финитных форм глагола *выступать* в Уппсальском корпусе равняется 86, а в Тюбингенском – 32. Получается, что для глагола *выступать* формы Ж. р. Прош. вр. составляют приблизительно 6% от всех финитных форм в обоих корпусах. Данные по всем 213 глаголам были объединены, и, неожиданно для нас, в двух словарях независимо друг от друга получилась похожая картина.

	Инфинитив	Настоящее время					
		Ед. ч.			Мн. ч.		
		1 л.	2 л.	3 л.	1 л.	2 л.	3 л.
Уппсальский корпус	18.8%	2.5%	1.5%	17.5%	1.7%	0.8%	7.8%
Тюбингенский корпус	20.4%	6.4%	1.7%	19.0%	2.5%	2.4%	7.5%

	Прошедшее время				Императив	
	М. р.	Ж. р.	Ср. р.	Мн. ч.	Ед. ч.	Мн. ч.
Уппсальский корпус	20.9%	10.1%	6.0%	10.7%	1.1%	0.5%
Тюбингенский корпус	14.6%	7.2%	5.3%	10.8%	0.6%	1.0%

Из таблиц видно, что количество тех или иных форм в наших корпусах может несколько различаться. Так, например, в разговорных текстах, как и следовало ожидать, больше форм 2 л. (так как в корпус вошли не дружеские беседы, а интервью из журналов

и газет, это заметно только по количеству форм 2 л. Мн. ч.). С другой стороны, в Тюбингенском корпусе меньше форм Прош. вр. Однако соотношение частотностей разных форм в двух корпусах во многом совпадает. Так как мы опирались на данные очень качественного корпуса письменных текстов и "проверяли" их на материале корпуса разговорных текстов, мы можем заключить, что получили соотношение между частотностями глагольных форм, близкое к реальному. Так, мы можем быть уверены, что в Наст. вр. наиболее частотной является форма 3 л. Ед. ч., а самыми редкими – форма 1 л. Мн. ч. и обе формы 2 л.

Для нашего эксперимента мы выбрали форму 2 л. Ед. ч. (одну из самых редких), форму 3 л. Ед. ч. (наиболее частотную из личных форм Наст. вр.) и инфинитив. В принципе, по частотности инфинитив почти совпадает формой с 3 л. Ед. ч. Однако существует гипотеза, что у инфинитива среди всех прочих форм должен быть особый статус. Психолингвистическая сущность этого "особого статуса" пока представляется довольно туманной. Не вдаваясь в теоретические дебри, мы решили проверить, удастся ли нам получить какие-то экспериментальные данные, подтверждающие тот факт, что инфинитив – это "особенная" форма.

В целом суть задуманного нами эксперимента была такова. Набрав две группы глаголов, относительно частотных и очень редких, мы собирались сравнить среднее время реакции на формы 2 л. и 3 л. Ед. ч. и на инфинитив в каждой группе. Понятно, что у каждого конкретного глагола соотношение между частотами этих форм отличается от вычисленного нами среднего соотношения. Однако, судя по обработанным нами 213 глаголам, форма 2 л. Ед. ч. всегда менее частотна, чем форма 3 л. Ед. ч. (кроме глаголов типа *видеть, слышать, помнить*). Сравнить 3 л. Ед. ч. и инфинитив сложнее. Инфинитив иногда частотнее формы 3 л. Ед. ч., а иногда реже. На большом множестве глаголов различия между ними нивелируются (как в случае с нашими 213 словами).

С точки зрения соотношения между 2 и 3 л. существовало три варианта развития событий. Во-первых, могло оказаться, что, с учетом всех прочих факторов, влияющих на время реакции, форма 3 л. Ед. ч. распознается быстрее формы 2 л. Ед. ч. и эта тенденция статистически значима в обеих группах. Во-вторых, такая тенденция могла наблюдаться только в более частотной группе, а в-третьих, и вовсе не проявиться. Подобным образом обстояло дело и с инфинитивом: либо он будет, несмотря на нивелированную разницу в

частотностях, значимо отличаться по времени реакции от формы 3 л. Ед. ч. в обеих группах, либо только в одной группе, либо между ними вообще не будет никакой статистически подтвержденной разницы. У каждого из этих вариантов может быть несколько объяснений в рамках разных теоретических подходов. Мы решили, что не стоит уделять здесь место таким сомнительным рассуждениям, как потенциальные объяснения всех неподтвердившихся сценариев. Обсуждение подтвердившегося варианта, то есть анализ данных, которые мы получили, помещено в подглаву "Результаты эксперимента".

4. Экспериментальный дизайн.

Эксперимент, спланированный нами, выглядел следующим образом. Как и в эксперименте Алегре и Гордона, испытуемым на экране компьютера предъявлялись слова и квазислова. Для предъявления стимулов использовалась программа Presentation, работающая на компьютерах типа РС. К ней прилагается собственный язык программирования, при помощи которого можно описать необходимую для конкретного эксперимента последовательность действий, например, сортировку стимулов в случайном порядке, вывод стимула на экран (или его озвучивание), измерение времени реакции. Подробное описание Presentation, ее достоинств и недостатков (которые, заметим в скобках, нас не коснулись) можно найти на ее официальном сайте www.neurobehavioralsciences.com. Главное достоинство программы заключается в том, что она позволяет измерять время реакции с точностью до 0,1 мс. Интересно отметить, что на компьютерах типа Macintosh такая точность измерений становится возможной только при наличии специального устройства, которое присоединяется к компьютеру. Программирование данного эксперимента было осуществлено совместно со студентом биологического факультета СПбГУ Даниилом Кислюком.

Испытуемым надо было ответить, видели ли они слово или квазислово, нажав на одну из двух клавиш на клавиатуре. Мы выбрали клавиши F и J и меняли их местами. Например, у первого испытуемого в первой части эксперимента J обозначала квазислово, во второй – слово, а в третьей – снова квазислово. У второго испытуемого все было наоборот. Мы просили испытуемых перед началом эксперимента положить указательные пальцы правой и левой руки на клавиши F и J и в дальнейшем уже не двигать руками. Это было сделано для того, чтобы сократить и усреднить время, затрачиваемое на

программирование и выполнение движений пальцев. Ведь это время является неотъемлемой частью общего времени реакции, и наша цель – свести влияние этого компонента к минимуму. Так как правая и левая рука неравноправны, мы постоянно чередовали клавиши, хотя в нашем случае это, судя по результатам эксперимента, было излишней предосторожностью.

Новый стимул каждый раз выбирался случайным образом из всего множества стимулов, которые еще не предъявлялись данному испытуемому, и появлялся на черном экране на 500 мс. Стимулы были набраны белым шрифтом (24-м кеглем) и возникали в центре экрана. После исчезновения стимула испытуемый видел просто черный экран. Время реакции отсчитывалось с момента появления стимула на экране (испытуемый мог нажать на клавишу, когда стимул еще не исчез, или уже после его исчезновения). После того, как испытуемый нажимал на клавишу или по прошествии 6000 мс (на случай, если человек растеряется и не сможет ответить), инициировался межстимульный интервал. Интервалы между стимулами произвольно варьировали от 700 до 1400 мс (то есть программа каждый раз случайным образом выбирала из множества чисел: 700, 750, 800 и так далее до 1400).

Так как человеку, не связанному с изучением времени реакции, все эти цифры ничего не скажут или даже покажутся абсурдными (может ли человек прочитать что-то на экране за полсекунды?), мы решили поместить здесь небольшой комментарий. Алегре и Гордон, к эксперименту которых мы все время обращаемся, использовали другую тактику. У них каждое слово появлялось на экране и не исчезало, пока испытуемый не нажмет на клавишу. Если испытуемый не реагировал в течении 1500 мс, на экране появлялось сообщение "слишком медленно". Соответственно, ответы, данные по прошествии 1500 мс, не учитывались. После того, как человек нажимал на клавишу, инициировался межстимульный интервал: на 1500 мс на экране появлялась так называемая "точка фиксации", роль которой выполняла звездочка.

Судя по ряду более поздних работ, включающих эксперименты такого рода, созданный нами экспериментальный дизайн является более "современным" и, как нам кажется, более правильным. Во-первых, цель эксперимента – заставить человека реагировать настолько быстро, насколько он может, чтобы добиться непосредственной реакции. Таким образом мы минимизируем раздумья и различные вторичные реакции на

стимулы (например, ассоциативные: это квазислово смешное, это слово у меня связано со словарными диктантами и т.п.) – ведь все это нас не интересует, а на времени реакции отражается. У Алегре и Гордона темп задается искусственно: если человек отвечает слишком медленно, то получает предупреждение (а если просто медленно, например, через 1100 мс вместо стандартных 500-700, то это никак не корректируется). В нашем эксперименте темп задается естественным образом: так как стимул быстро исчезает, человек отвечает сразу же, пока он еще помнит, что было на экране. При этом человеку вполне достаточно полсекунды, чтобы прочитать на экране слово длиной в 6-10 букв (это подтвердилось и в ходе нашего эксперимента – большая часть испытуемых ошибалась довольно редко). В англоязычных экспериментах часто используется более короткое время предъявления стимулов, чем в нашем, но это связано с тем, что английские слова в среднем короче.

Далее, у Алегре и Гордона не изменяется межстимульный интервал, в то время как в большинстве современных работ он варьирует. Это делается для того, чтобы избежать эффекта ожидания: в течение эксперимента человек привыкает к тому, что каждый новый стимул появляется через определенное время, и более ранние стимулы оказываются в неравных условиях с более поздними. Правильнее, чтобы все стимулы появлялись неожиданно. Заметим, что в англоязычных работах межстимульные интервалы обычно меньше, чем у нас, но это, опять же, связано с тем, что английские слова короче и эксперимент в целом идет быстрее. В некоторых работах (см., например, классическую для нейролингвистики статью Embick et al. 2001) используются крайне короткие межстимульные интервалы, чтобы окончательно сбить испытуемых с толку. Нам кажется, что это неправильно, так как тогда после появления следующего стимула человек некоторое время не может на нем сосредоточиться. Что же касается точки фиксации, то она обычно используется для того, чтобы человек не отводил глаза от центра экрана. В каких-то экспериментах (например, когда стимулы появляются в разных частях экрана и учитывается время, которое необходимо, чтобы перевести взгляд) без точки фиксации не обойтись. В экспериментах, подобных нашему, в ней нет особого смысла, а необходимость все время глядеть на нее довольно утомительна.

Единственная явная ошибка, допущенная нами при программировании эксперимента, – это то, что в случае, если человек не отвечает, межстимульный интервал

инициировался только через 6000 мс, то есть через 6 секунд. На самом деле на этом месте должно было стоять число 1500. Дело в том, что во многих подобных экспериментах ответы, в которых время реакции превышает 1500 мс, все равно отбрасываются – это слишком долго для непосредственной реакции (ср., например, Alegre and Gordon 1999, Embick et al. 2001). Так что логично было бы переходить к межстимульному интервалу уже по прошествии 1500 мс. Однако, пока мы отлаживали программу, мы заменили 1500 мс на 6000, и забыли потом поменять это число обратно. На результатах эта ошибка никак не отразилась, так как все случаи, где время реакции превышало 1500 мс, все равно не учитывались (мы использовали более сложный механизм, чем отбрасывание всех реакций после какого-то порога; об этом см. подглаву "Статистическая обработка данных").

В эксперименте приняло участие 28 русскоговорящих людей: 16 женщин и 12 мужчин в возрасте от 15 до 50 лет. Заметим, что результаты эксперимента показали, что ни возраст, ни пол испытуемых не стали хоть сколько-нибудь заметными факторами. Данные, полученные от трех человек, пришлось исключить (см. подглаву "Статистическая обработка данных"), так что в результате осталось 25 испытуемых.

В качестве стимулов были использованы следующие слова и квазислова. Во-первых, глаголы АЙ, А, Е, И и ОВА классов. В каждом классе было взято 5 глаголов с частотностью около 30 на миллион, 5 очень редких и 10 квазиглаголов. Каждый глагол появился в эксперименте в трех формах: инфинитив, 2 л. Ед. ч. и 3 л. Ед. ч. Относительно частотные глаголы отбирались по словарю Лённгрена. Как мы уже отмечали выше, у нас есть основания не доверять словарным частотностям. Однако в данном случае нас, в отличие от Алегре и Гордона, интересовала не точная цифра, а приблизительный диапазон частот, о котором словарь способен дать представление. Очень редкие глаголы отбирались интуитивно и затем "проверялись" на нескольких носителях русского языка (мы просто спрашивали, кажутся ли им эти слова крайне малоупотребительными и редкими). Безусловно, это кустарный метод отбора, однако на данный момент это единственный способ работать с низкочастотными русскими словами. Кроме того, нас не пугал тот факт, что у одного из выбранных нами слов могла оказаться частотность 1 на миллион, а у другого – в два раза меньше, 0,5. Главное, что все они наверняка были существенно более редкими, чем глаголы первой группы (все глаголы с частотностью ~30 на миллион интуитивно воспринимаются как частотные). Квазиглаголы создавались

путем замены в основах исходных глаголов нескольких звуков, несущественных с точки зрения возможных чередований.

Слова с частотностью ~30 на миллион были выбраны по двум причинам. Во-первых, только в этой частотной группе удалось найти достаточное количество глаголов, которые можно было сбалансировать по всем необходимым параметрам (длина, вид, класс, отсутствие приставки). Во-вторых, частотность формы 3 л. Ед. ч. и инфинитива у таких глаголов будет приблизительно равна 6 на миллион (~20% от частоты гнезда), а частотность формы 2 л. Ед. ч. будет заведомо меньше 1 на миллион. Алегре и Гордон, как мы помним, показали в своем эксперименте, что у англоговорящих людей в памяти хранятся формы с частотностью большей или равной приблизительно 6 на миллион. Обсуждая теоретические проблемы, связанные с противопоставлением хранения в памяти и морфологического анализа при восприятии, мы уже говорили о том, что соотношение между такими языками, как английский и русский, в этой области (как и во многих других) пока не известно науке. Мы решили, что, раз нам представилась такая возможность, будет очень интересно сравнить их.

Кроме того, чтобы отвлечь испытуемых, были отобраны 30 реальных существительных (15 с частотностью ~30 на миллион и 15 очень редких) и 30 квазисуществительных, которые также фигурировали в трех формах: Им. п. Ед. ч., Тв. п. Ед. ч. и Дат. п. Мн. ч. Число существительных мужского и женского рода во всех группах было приблизительно равным. Таким образом, в эксперименте было всего 480 стимулов длиной от 6 до 10 букв, и он длился около 20 минут. Перед началом эксперимента испытуемые могли потренироваться на 12 пробных словах.

Чтобы две формы одного и того же слова не встретились в эксперименте друг за другом (в этом случае благодаря эффекту прайминга распознавание второй формы будет ускорено), мы поделили все стимулы на три группы. В каждую из них вошло по одной форме каждого слова и квазислова. Эти группы были уравновешены по всем параметрам (количество реальных и квазислов, глаголов и существительных, частых и редких слов, количество глаголов разных классов, существительных мужского и женского рода, количество разных глагольных и именных форм). Далее мы создали 6 вариантов эксперимента, в каждом из которых эти три группы следовали друг за другом в разном порядке. Таким образом, появление форм одного и того же слова было по возможности

разнесено во времени, и каждому следующему испытуемому формы предъявлялись в новой последовательности. Внутри группы, как уже было сказано выше, стимулы каждый раз сортировались в произвольном порядке.

5. Статистическая обработка данных.

Непосредственным результатом нашего эксперимента были 13440 реакций, полученных от 28 человек в ответ на предъявленные нами 480 стимулов. Каждый "ответ" включал в себя два компонента: время реакции и ее правильность (правильно ли человек определил, слово он видел или нет). Время реакции в дальнейшем обозначается в таблицах и графиках как RT (от *reaction time*), а правильность – как асс (от *accuracy*). На время реакции на каждую конкретную форму влияет множество факторов. Часть из них объективные, которые мы можем учесть и проанализировать (их список помещен ниже), другие совершенно случайны и не поддаются никакому учету. Любая посторонняя мысль может отвлечь человека и существенно замедлить реакцию на один или несколько стимулов.

Как же обрабатываются данные? Влияние объективных факторов на время реакции вычисляется при помощи специального статистического метода – дисперсионного анализа. Случайными факторами приходится пренебречь, утешаясь тем, что их влиянию мог с равной вероятностью подвергнуться любой стимул. То, что многие эксперименты с использованием времени реакции дали статистически значимые результаты, доказывает, что отбрасывать случайные факторы оправдано.

Дисперсионный анализ используется сейчас в подавляющем большинстве работ, в которых требуется изучение многофакторных планов. Поэтому в статьях обычно просто приводятся его результаты, без каких-либо объяснений. Мы также следуем этой традиции. Однако, насколько нам известно, в отечественных работах дисперсионный анализ пока применяется довольно редко, причем даже в тех областях, которые по уровню статистической обработки данных должны сильно опережать лингвистику¹³. Поэтому мы

¹³ Эта проблема обсуждается, например, в опубликованной в Интернете статье В. Леонова "Долгое прощание с лысенковщиной". Статья основана на фактах, полученных автором при анализе нескольких сот диссертаций и монографий, а также более 1500 статей в области экспериментальной биомедицины (<http://www.biometrica.tomsk.ru/lis/index.htm>).

решили посвятить дисперсионному анализу одно из приложений, в котором вводятся основные связанные с ним понятия и объясняются его методы.

Вот список объективных факторов для нашего эксперимента:

1. Форма слова (form). Это тот фактор, влияние которого, мы, собственно, и хотим определить путем данного эксперимента.

2. Количество повторений (обозначается буквосочтанием гер от *repetition*). Хотя мы сделали многое, чтобы подавить эффект прайминга, когда слово появлялось перед испытуемыми во второй раз (в новой форме), оно распознавалось быстрее, чем в первый. В третий раз время реакции также сокращалось по сравнению с вторым. Конечно, речь идет не о продолжающейся активации (lasting activation) данной лексемы в стогомом смысле слова, ведь ее формы входили в разные части эксперимента и между ними были интервалы в несколько минут. Однако, как мы знаем из других исследований, какая-то "остаточная" активация может сохраняться у слова еще очень долго после того, как человек встретил его в тексте или в речи. Именно с ее действием мы и столкнулись. Этот фактор противодействует интересующему нас фактору: быстрее распознается не какая-то определенная форма, а та, что появилась позже.

3. Длина слова (bukv). Длина слов, предъявляемых визуально, измеряется количеством букв. Это не настолько очевидно, так как, например, количество слогов также должно играть роль при распознавании слова. Однако обычно, чтобы не усложнять картину, это игнорируется. Чем длиннее слово, тем медленнее оно прочитывается и распознается. Важно понимать, что это фактор, существенный и для сравнения форм внутри одного слова. Так, в русском языке форма 2 л. Ед. ч. всегда на одну букву длиннее, чем форма 3 л. Соотношение длины форм настоящего времени и форм инфинитива варьирует от класса к классу: в классе АЙ форма 3 л. Ед. ч. совпадает по длине с инфинитивом, в классах А, И и Е меньше на одну букву, а в классе ОВА – на две. Этот фактор также противодействует фактору "форма слова": медленнее читается не какая-то определенная форма, а та, что длиннее.

4. Частотность (freq). Слова из редкой группы распознаются медленнее, чем из более частотной. Прогноз относительно квазислов менее очевиден. Иногда они распознаются медленнее реальных редких, а иногда быстрее редких, но медленнее частотных (если в них есть, например, бросающиеся в глаза сочетания букв, заведомо

невозможные в данном языке). Существует ряд экспериментов, специально посвященных этому явлению. Действие этого фактора не "смешивается" с действием того, который мы изучаем. По сути, фактор частотности задает три группы стимулов, внутри которых мы будем анализировать и сравнивать действие фактора "форма слова".

5. Глагольный класс (class). В эксперименте могла бы проявиться какая-то особая зависимость времени реакции от глагольного класса. Однако мы сразу скажем, что этот фактор не оказался статистически значимым и не оказывал на время реакции никакого явного влияния.

Еще одним фактором можно считать личность испытуемого. Среднее время реакции очень сильно варьирует в зависимости от человека. Однако этот фактор не принято включать в факторную модель для дисперсионного анализа. Ведь он не противодействует ни одному из изучаемых экспериментаторами факторов, о его существовании всем известно а priori, так что статистическое подтверждение его значимости никого не интересует.

Статистическую обработку данных мы начали с того, что посчитали для каждого испытуемого среднее время реакции и процент правильных ответов. Все участники эксперимента здесь и далее обозначаются двумя инициалами. Мы решили установить пороговые значения, после которых ответы данного испытуемого следует исключить: среднее время реакции больше 1000 мс или меньше 80% правильных нажатий на клавишу (это достаточно условные числа, мы выбрали их исходя из общего представления о нормальном времени реакции и неслучайном количестве верных ответов).

№	Инициалы	Среднее время реакции	Процент правильных ответов
1	AD	806.9	94.6%
2	AR	1091.7	89.0%
3	AS	666.1	90.6%
4	ASh	1196.7	94.6%
5	AZ	837.6	80.4%
6	GJ	682.2	88.8%
7	IG	860.2	94.2%
8	IN	601.8	91.9%
9	JR	640.8	87.1%
10	KF	659.1	90.8%
11	KS	705.0	95.0%

12	KSo	641.1	93.3%
13	KV	963.7	69.6%
14	LS	825.8	90.8%
15	MK	633.0	89.8%
16	MS	642.8	86.7%
17	MV	706.0	93.3%
18	MX	683.0	96.0%
19	ND	897.9	90.6%
20	NKa	744.1	90.6%
21	NP	585.3	80.6%
22	ON	904.3	94.4%
23	PE	643.9	90.4%
24	PP	550.2	87.3%
25	PS	681.2	85.6%
26	RF	806.2	96.5%
27	SL	873.9	95.0%
28	TP	799.3	94.8%

Как видно из этой таблицы, данные от трех участников: ASh, AR и KV пришлось исключить, и у нас осталось 25 испытуемых. После этого обычно принято выбрасывать все ответы после какого-то порогового значения, скажем, после 1500 мс (ср., например, Alegre and Gordon 1999, Embick et al. 2001). Ведь очевидно, что время реакции не могло измениться так сильно под влиянием какого-то из изучаемых нами объективных факторов, так как ответ слишком сильно выбивается из общего ряда. Значит, на эту реакцию повлияли какие-то случайные факторы, иначе говоря, человек отвлекся, или стимул вызвал у него какие-то посторонние ассоциации (удивил, рассмешил, о чем-то напомнил).

Однако нам такой подход не кажется вполне оправданным. Как можно заключить из приведенной выше таблицы, у всех участников было разное среднее время реакции. Так, например, у "самого быстрого" испытуемого PP среднее время реакции – 550 мс. Интуиция подсказывает, что для него не только 1500, а и 900 мс – это аномально долгая реакция. С другой стороны, у "самого медленного" испытуемого ON 900 мс – это среднее время реакции. Как дать подобным интуитивным соображениям научную основу и установить свое пороговое значение для каждого из испытуемых? Мы воспользовались для этого такой несложной статистической процедурой, как определение экстремальных значений, или выбросов. Комментарий по поводу того, как определяются экстремальные значения и почему их можно и нужно исключать из рассмотрения, мы поместили в том же

приложении, где обосновывается применение в данной работе дисперсионного анализа. Здесь же, чтобы не перегружать текст статистическими отступлениями, мы просто приведем таблицу, из которой явствует, сколько ответов было исключено и на каких основаниях:

Испытуемые	После исключения тех случаев, когда не было получено вообще никакой реакции ¹⁴		После исключения тех случаев, когда время реакции было слишком маленьким		После исключения выбросов		
	Всего	Гл. и квазигл.	Всего	Гл. и квазигл.	Выбросы > или =	Всего	Гл. и квазигл.
AD	480	300	480	300	1119 мс	461	287
AS	480	300	480	300	935 мс	461	289
AZ	478	299	478	299	1267 мс	450	284
GJ	479	299	479	299	967 мс	453	284
IG	480	300	479	299	1162 мс	448	277
IN	479	299	479	299	898 мс	470	294
JR	480	300	480	300	928 мс	474	296
KF	480	300	480	300	918 мс	455	285
KS	479	299	479	299	1216 мс	474	296
KSo	480	300	480	300	933 мс	467	294
LS	480	300	480	300	1159 мс	457	285
MK	480	300	480	300	924 мс	475	297
MS	480	300	479	300	901 мс	463	291
MV	480	300	480	300	1013 мс	466	291
MX	480	300	480	300	1025 мс	465	291
ND	479	299	479	299	1312 мс	438	268
NKa	480	300	480	300	1065 мс	471	298
NP	475	296	475	296	795 мс	453	281
ON	480	300	480	300	1455 мс	470	296
PE	480	300	480	300	927 мс	471	295
PP	479	300	479	300	794 мс	448	280
PS	478	298	477	297	1078 мс	460	284
RF	480	300	480	300	1352 мс	465	290
SL	480	300	480	300	1241 мс	468	291
TP	479	299	479	299	1242 мс	462	286

Из-за "слишком маленького времени реакции" было исключено всего три ответа. Во всех трех время реакции было меньше 82 мс, в то время как все остальных случаях –

¹⁴ То есть испытуемый так и не нажал ни на одну из клавиш.

больше или равно 264 мс (причем через такое короткое время начинали отвечать только те несколько участников, у которых было необычно быстрое время реакции). Заметим, что мы оставили те ответы, в которых испытуемые нажимали не на ту клавишу. Судя по тому, что участники рассказывали нам после эксперимента, в большинстве случаев они ошибались не потому, что не успевали прочесть стимул, а потому, что путали клавиши. То есть решение о том, что пора нажимать на клавишу, принималось после того как стимул был опознан, но при этом сделать это могла, например, та же рука, что и в прошлом ответе. Вообще, в отличие от чтения и понимания слов, нажатие на кнопки то правой, то левой рукой – совершенно неестественное и непривычное для человека занятие. Мы знаем, что такая позиция может вызвать возражения. Поэтому мы обсчитывали все данные также и без неправильных ответов. И только убедившись, что это никак не влияет на конечный результат, отбросили этот вариант расчетов.

Те ответы, которые решено было оставить, мы подвергли дисперсионному анализу при помощи SPSS – на сегодняшний день наиболее мощной и самой популярной среди исследователей различных направлений статистической программы. Процедура определения экстремальных значений также была осуществлена посредством этой программы. Результаты, так как это, по сути, и есть итог нашего эксперимента, приводятся в следующей подглаве.

6. Результаты эксперимента.

Дисперсионный анализ всего множества ответов показал, что из пяти объективных факторов, присутствовавших в нашем эксперименте, (количество повторений, длина слов, частотность, форма слов и глагольный класс) значимыми оказались четыре. Так как глагольный класс не стал значимым фактором, мы исключили его из таблиц дисперсионного анализа:

ANOVA: все стимулы		Sum of Squares	df	Mean Square	F	Sig.	
RT	Main Effects	(Combined)	14997101.3	10	1499710.1	66.21	0.000
		REP	1955942.0	2	977971.0	43.17	0.000
		BUKV	298343.1	4	74585.8	3.29	0.011
		FORM	194340.9	2	97170.5	4.29	0.014
		FREQ	12212069.6	2	6106034.8	269.56	0.000

Model	14997101.3	10	1499710.1	66.21	0.000
Residual	163069440.3	7199	22651.7		
Total	178066541.6	7209	24700.6		

Так как таблица дисперсионного анализа (или ANOVA) показывает только важность или неважность какого-то фактора, для наглядности мы приводим также таблицы со средними значениями времени реакции:

FREQ	Среднее время реакции	Кол-во
Квазиглаголы	725.1 мс	3572
Редкие глаголы	700.3 мс	1791
Глаголы с частотностью ~30 на млн.	624.4 мс	1847
Всего	693.1 мс	7210

REP	Среднее время реакции	Кол-во
1 раз	713.9 мс	2362
2 раз	692.3 мс	2411
3 раз	673.9 мс	2437
Всего	693.1 мс	7210

BUKV	Среднее время реакции	Кол-во
6 букв	691.3 мс	650
7 букв	684.4 мс	1861
8 букв	692.0 мс	3281
9 букв	705.4 мс	1327
10 букв	746.8 мс	91
Всего	693.1 мс	7210

FORM	Среднее время реакции	Кол-во
2 л. Ед. ч.	701.4 мс	2393
3 л. Ед. ч.	691.2 мс	2416
Инфинитив	686.9 мс	2401
Всего	693.1 мс	7210

Итак, мы видим, что все факторы оказались статистически значимыми на очень высоком уровне, то есть все они, несомненно, влияли на время реакции. Однако среди них можно установить иерархию (исходя из значений F). Самым важным фактором оказалась частотность, затем – количество повторений. Длина слова более важна, чем форма. В

случае с длиной мы даже можем видеть по средним значениям времени реакции, что зависимость нарушается для шестибуквенных слов. Форма слова оказалась статистически значимым фактором, однако наименее важным среди всех. Заметим, что не только 3 л. Ед. ч. распознается быстрее, чем 2 л., но и инфинитив быстрее, чем 3 л., однако разница между ними довольно невелика.

Теперь рассмотрим квазиглаголы и две группы реальных глаголов по отдельности.

ANOVA: квазиглаголы			Sum of Squares	df	Mean Square	F	Sig.
RT	Main Effects	(Combined)	1557329.9	8	194666.2	8.16	0.000
		REP	1383586.2	2	691793.1	28.91	0.000
		BUKV	172522.5	4	43130.6	1.80	0.126
		FORM	5335.7	2	2667.9	0.11	0.895
	Model		1557329.9	8	194666.2	8.14	0.000
	Residual		85255634.1	3563	23928.0		
	Total		86812964.1	3571	24310.5		

ANOVA: редкие глаголы			Sum of Squares	df	Mean Square	F	Sig.
RT	Main Effects	(Combined)	1110927.3	8	138865.9	5.34	0.000
		REP	726316.3	2	363158.1	13.97	0.000
		BUKV	306477.8	4	76619.4	2.95	0.019
		FORM	101046.4	2	50523.2	1.94	0.143
	Model		1110927.3	8	138865.9	5.34	0.000
	Residual		46317350.7	1782	25991.8		
	Total		47428278.0	1790	26496.2		

ANOVA: глаголы с частотностью ~30 на млн.			Sum of Squares	df	Mean Square	F	Sig.
RT	Main Effects	(Combined)	694509.7	7	99215.7	5.95	0.000
		REP	102865.2	2	51432.6	3.09	0.046
		BUKV	124428.0	3	41476.0	2.49	0.059
		FORM	441256.2	2	220628.1	13.24	0.000
	Model		694509.7	7	99215.7	5.95	0.000
	Residual		30649945.0	1839	16666.6		
	Total		31344454.7	1846	16979.7		

Для квазиглаголов важным фактором оказывается только количество повторений. Фактор длины слова не достигает статистической значимости, а форма слова не играет вообще никакой роли. Для редких глаголов статистически значимыми оказываются количество повторений ($p < 0,001$) и, в меньшей степени, длина слова ($p = 0,019$). Влияние глагольной формы больше, чем у квазиглаголов, но все же оно статистически незначимо ($p = 0,143$). В группе глаголов с частотностью около 30 на миллион количество повторений и длина слов влияют на время реакции, однако меньше, чем у редких глаголов ($p = 0,046$ и $p = 0,059$). Наиболее важным фактором, влияние которого подтвердилось на очень высоком уровне значимости ($p < 0,001$), оказалась форма глагола. Видимо, именно за счет действия этого фактора по сравнению с группой редких глаголов снизилось значение количества повторений и длины слова (эти факторы противодействуют друг другу).

Интерпретируя полученные нами данные, можно прийти к выводу, что некоторые формы глаголов с частотностью около 30 на миллион хранятся в памяти. Так как частота гнезда для форм одного и того же слова совпадает по определению, тот факт, что они распознаются с разной скоростью, указывает на то, что при восприятии в ментальном лексиконе форма ищется целиком. Наши данные также подтверждают особый статус инфинитива: приблизительно совпадая по частоте с формой 3 л. Ед. ч., он распознается значительно быстрее. Уточняя, какие именно из использованных в нашем эксперименте форм хранятся в памяти, мы можем предположить, что это форма 3 л. Ед. ч. и инфинитив. Для глаголов с частотностью около 30 на миллион собственная частотность этих форм составляет 5-6 на миллион. Таким образом, они находятся примерно в том же диапазоне, что и формы английских глаголов, которые, согласно экспериментам Алегре и Гордона, хранятся в памяти. Собственная частотность формы 2 л. для глаголов этой группы составляет примерно 0,4 на миллион. Маловероятно, что такие формы могут храниться в памяти. Тогда тот факт, что они распознаются медленнее других, объясняется тем, что они не находятся в ментальном лексиконе целиком, а подвергаются морфологическому анализу. Такая модель косвенно подтверждается тем, что мы не нашли существенного различия между формами редких глаголов. Ведь если бы оно существовало, разница во времени реакции на формы 2 и 3 л. Ед. ч. могла бы быть объяснена тем, что обе формы

подвергаются анализу, но флексия 3 л. намного частотнее, и потому оно распознается быстрее 2 л.

IV. База данных по русскому глаголу на основе "Грамматического словаря русского языка".

1. Описание базы данных.

Эта глава посвящена еще одной части проекта по экспериментальному изучению русской глагольной морфологии – созданию базы данных по русскому глаголу на основе "Грамматического словаря русского языка". В ряду прочих эта часть проекта стоит особняком, так как в данном случае речь шла не о проведении экспериментов, а о получении важной для них информации о системе русских глагольных классов.

Рассмотрев приведенные в этой работе экспериментальные данные, можно убедиться в том, что при планировании подобных экспериментов и при обработке полученного материала наиболее важными параметрами являются: объем глагольных классов (type frequency), сложность парадигмы и различные другие характеристики глагольных классов, делающие их "узнаваемыми" (так называемые морфологические подсказки, акцентные парадигмы и т.д.). Кроме того, огромную роль играют частотности отдельных глаголов (token frequency). Последние стоят в этом ряду особняком, так как, в отличие от прочих перечисленных параметров, не выявляются при анализе системы глагольных классов, а вычисляются на основе специально составленного корпуса текстов. Три первых параметра, напротив, можно получить, проанализировав в отрыве от текстов все существующие глаголы и различные особенности их спряжения. Для русского языка подобные расчеты были сделаны лишь отчасти, и мы решили заполнить эту лауну. Полученные нами сведения уже были использованы в рамках нашего проекта и могут в дальнейшем оказаться полезными и для других исследователей.

Для решения поставленных задач нами была создана база данных по русскому глаголу. За основу был взят "Грамматический словарь русского языка" (сост. А. А. Зализняк). В словаре Зализняка содержится более 27 тысяч глаголов, для которых указаны все словоизменительные характеристики. Сам А. А. Зализняк описал концепцию своего словаря следующим образом: "Полный грамматический словарь должен для каждого входящего в него слова указывать все те его характеристики, которые существенны для построения грамматически правильных фраз, содержащих данное слово. Таковы, во-первых, словоизменительные, во-вторых, синтаксические характеристики слова.

Настоящий словарь отражает полно только словоизменительные характеристики слова – отсюда его подзаголовок. Из синтаксических характеристик слова даются лишь немногие – главным образом те, которые наиболее тесно связаны со словоизменением¹⁵.

По поводу словника Зализняк пишет: "Словарь имеет нормативную ориентацию, т.е. он указывает для отдельного слова не все реально встречающиеся способы образования его форм, а лишь те, которые соответствуют современной литературной норме. Исключение составляют формы, помеченные как просторечные (введение в словарь некоторого числа таких форм продиктовано системными морфологическими соображениями). Помета "простореч." выступает в таких случаях в роли предупредительной. Разумеется, нормативность не исключает наличия вариантов: во многих случаях признаются допустимыми два или более варианта форм"¹⁶. Основу словника "Грамматического словаря русского языка" составил "Орфографический словарь"¹⁷. В спорных случаях привлекались толковые словари. Некоторое количество слов было добавлено из словаря-справочника "Новые слова и значения"¹⁸, охватывающего публицистику и художественную литературу 60-х годов, а также из более поздней прессы. Таким образом, большое число неологизмов, появившихся в последние десятилетия, в словарь не вошло. Можно, однако, утешаться тем, что новые глаголы намного более редки, чем новые существительные или прилагательные. Кроме того, мы не ставим себе задачу учесть все без исключения русские глаголы. Это нереально уже потому, что никакой словарь не может включить все возможные приставочные глаголы и производные глаголы несовершенного вида. Никогда нельзя с уверенностью сказать, имеет ли такой глагол сколько-нибудь значительную частотность, или, хотя каждый из нас теоретически мог бы употребить его, на деле этот глагол встречается крайне редко. Зализняк, упоминая

¹⁵ Зализняк 1987. С. 3

¹⁶ Зализняк 1987. С. 9

¹⁷ Орфографический словарь русского языка. Изд. 6-е. Под ред. С.Г. Бархударова, С.И. Ожегова и А.Б. Шапиро, М., 1965.

¹⁸ Новые слова и значения. Словарь справочник по материалам прессы и литературы 60-х годов. Под ред. Н.З. Котеловой и Ю.С. Сорокина, М., 1971

в предисловии эту проблему, признается, что "во всех этих случаях принятое в настоящем словаре решение содержит некоторый элемент условности"¹⁹.

Именно по этой причине, создавая базу данных, мы не вносили никаких изменений в список глаголов, позаимствованный нами из словаря Зализняка, и не пытались пересмотреть грамматический комментарий в тех вопросах, которые имеют отношение к языковой норме. Структура базы данных сильно отличается от структуры словаря, но только за счет того, что одни и те же формы и особенности спряжения по-разному называются и группируются. Начав дополнять словарь, мы поставили бы себя перед необходимостью переработать весь словник, подкрепляя свои соображения большим количеством примеров или ссылкой на авторитетные источники. Это очень непростая задача не только для одного человека, но и для группы квалифицированных лексикографов. Поэтому, даже если в ряде случаев решение составителей словаря кажется нам необоснованным, мы все равно ничего не меняли. Исключения составляют только явные ошибки, которые мы по мере возможности исправляли. В связи с этим следует заметить, что и в самом "Грамматическом словаре", и при переведении исходного материала в электронный формат было допущено некоторое количество ошибок. Число их в нашей базе невелико и постоянно уменьшается, однако вряд ли когда-нибудь удастся искоренить их окончательно.

База данных предстает собой файл Microsoft Excel и устроена следующим образом. Каждый глагол занимает одну строку, и вся касающаяся его информация рассредоточена по столбцам. Ниже для иллюстрации приведены шесть глаголов с грамматическим комментарием (строки разрезаны, так как уместить их на листе целиком было невозможно):

WORD	VICEVERSA	ACCENT	ASPECT	TRANS	CLASS	#, ?	ALTERN
запереться	ясьтерепаз	6	св		/P		
запереться	ясьтерепаз	6	св		/P		
слабеть	ьтебалс	5	нсв	нп	ЕЙ		
пожалеть	ьтелажоп	6	св		ЕЙ		
номеровать	ьтаворемон	8	нсв		ОВА		
обжечься	ясьчежбо	4	св		Г-К		Г
поздоровиться	ясьтивородзоп	7	св		И		

¹⁹ Зализняк 1987. С. 8

WORD
запереться
запереться
слабеть
пожалеть
номеровать
обжечься
поздоровиться

PRAES	PAST	ASPECT PAIR	ПОМЕТЫ	*	■, □	ë, o
b	c"	\$II (запира'ться)	"1"			
b	a	\$II (запира'ться)				ë
a	a					
a	a				□	
a	a					
b	b	\$II (обжига'ться)		*		ë
a	a					

WORD
запереться
запереться
слабеть
пожалеть
номеровать
обжечься
поздоровиться

бзл., мнгр.	GRAMMAR
	<i>прош. м</i> заперся' // за'перся; <i>прич. прош.</i> за'першийся; <i>деепр. прош.</i> заперши'сь
	<i>буд.</i> обож гу'сь, -жётся, -гут'ся; <i>прош.</i> обжётся, обожгла'сь, обжётся
безл.	

WORD
запереться
запереться
слабеть
пожалеть
номеровать
обжечься
поздоровиться

IDIOMS	VARIANTS	MEANING & USAGE
		<i>закрыться</i>
		<i>заупрямиться</i>
	нумерова'ть	
кому'-л. не здоро'вится		

WORD	Глагол X.
VICEVERSA	"Перевернутый" глагол X, позволяющий сортировать слова в инверсионном алфавитном порядке. Такой способ принят в словаре Зализняка. Он очень удобен для таких словарей, так как вся грамматическая информация сосредоточена в русском языке в конце слова.
ACCENT	Ударение (на какую "букву" от начала оно падает). Дело в том, что указывать ударение при помощи знака " ' " в первой колонке было бы крайне неудобно. Ведь для этого пришлось бы ввести особые литеры. Это затруднило бы поиск и помешало бы правильной сортировке глаголов по алфавиту. Поэтому мы и прибегли к этому кустарному, но абсолютно эффективному способу. В скобках указывается вторичное ударение.
ASPECT	Вид глагола X: св, нсв, св-нсв (=двувидовой)

TRANS	Переходность глагола X. Ставится помета "нп", только если глагол непереходный во всех своих значениях. У возвратных глаголов такая помета отсутствует, так как они всегда непереходные
CLASS	Глагольный класс (по Одноосновной Системе Якобсона). Система Якобсона объясняется в Приложении 1. Таблицы, позволяющие сравнить ее с системой, принятой в Академической грамматике и в словаре Зализняка, находятся в Приложениях 2 и 3 соответственно.
#, ?	Сделанные мной пометы, позволяющие отследить, где классификация Якобсона расходится с классификацией Зализняка (помечены спорные случаи)
ALTERN	Конечный согласный основы. Указываются для тех глаголов, у которых его невозможно определить исходя из их внешнего вида и морфологического класса. Например, <i>мести</i> (класс Д-Т) – <i>мету</i> , но <i>вести</i> (класс Д-Т) – <i>веду</i> .
PRAES	Акцентная парадигма глагола X в настоящем/будущем времени (объяснения обозначений см. в Приложении 5).
PAST	Акцентная парадигма глагола X в прошедшем времени (объяснения см. в Приложении 5).
ASPECT PAIR	Глагол противоположного вида, составляющий вместе с глаголом X видовую пару. Если он образуется стандартным способом, указывается только номер (объяснения см. в Приложении 5).
POMETY	Цифровые пометы. Все пометы здесь и ниже соответствуют тем, которые приняты в словаре Зализняка
*	Помета * (в словаре Зализняка указывается после номера глагольного класса)
■, □	Пометы <input checked="" type="checkbox"/> и <input type="checkbox"/>
ě, o	Пометы <i>ě</i> и <i>o</i>
бзл., мнкр.	Пометы <i>бзл.</i> (глагол, безличный во всех своих значениях) и <i>мнкр.</i> (глагол, обозначающий многократное действие)
GRAMMAR	Словесный грамматический комментарий. Если спряжение глагола не определяется его классом и сопровождающими пометами, неочевидные формы указываются полностью (с ударениями) в этой графе
IDIOMS	В приведенных в этой графе идиомах слово употребляется в необычной форме или слово вообще не употребляется вне приведенных сочетаний
VARIANTS	Абсолютный синоним глагола X, если таковой имеется (варианты типа <i>номеровать</i> – <i>нумеровать</i> и пр.)
MEANING & USAGE	Значение глагола X. Указывается только тогда, когда есть два полных омонима или два глагола, различающихся только местом ударения, которые имеют разные словоизменительные характеристики или входят в разные видовые пары. Иногда

	значение может объясняться не описательно, а через возможные зависимые слова. Например, у глагола <i>побить</i> с парным <i>побивать</i> указано <i>врага, рекорд, посевы</i> , а у <i>побить</i> без парного глагола НСВ – <i>прочие значения</i> .
--	--

Значения той или иной пометы, не описанные в этой таблице, указаны в Приложении 5. Там же содержатся и прочие объяснения, касающиеся использованных в базе данных обозначений.

2. Сведения о глагольной системе русского языка, полученные при помощи базы данных

Первоначальная задача, послужившая поводом для создания базы данных, заключалась в том, чтобы узнать объем русских глагольных классов (type frequency) для эксперимента на порождение форм от квазиглаголов, описанного в первой главе. Раньше все, в том числе мы сами, черпали представления о системе русских глагольных классов из Академической грамматики. В ней перечислены только непроеизводные глаголы (бесприставочные немотивированные), входящие в некоторые классы. Получить представление об общем количестве глаголов в том или ином классе по Академической грамматике невозможно. Поэтому, говоря об объеме глагольных классов при анализе экспериментов, мы попросту делили их на продуктивные, которые все без исключения считали очень большими, и непродуктивные, для которых мы знали число входящих в них неприставочных глаголов. Когда, пользуясь базой данных, мы подсчитали объем глагольных классов, выяснилось, например, что продуктивные классы АЙ и ЕЙ, которые мы ранее, за неимением статистических данных, ставили в один ряд, отличаются по величине примерно в 18 раз.

название	количество глаголов	продуктивность
АЙ	11735	продуктивный
И	7174	продуктивный
ОВА	2815	продуктивный
НУ	1072	продуктивный
А	842	
ЕЙ	638	продуктивный
Е	326	
(НУ)	305	

Г-К	257	
ЖА	247	
n/s A	247	
Д-Т	202	
З-С	180	
(И)Й	163	
ОЙ	98	
АВАЙ	93	
/М-/Н	87	
/Р	80	
О	71	
Н	59	
В	50	
Й	45	
Й/М // НИМ	41	
Б	39	

Кроме того, существует ряд глаголов, для которых возможны формы, образованные по модели двух разных классов:

А // АЙ	108
А // n/s A	3
АЙ // А	18
ЕЙ // Е	3

Узнав объем глагольных классов, мы заинтересовались тем фактом, что, например, непродуктивный класс А превосходит продуктивный класс ЕЙ. Было бы крайне странно, думали мы, если бы в продуктивном классе было меньше бесприставочных глаголов, чем в непродуктивном. Здесь заметим, что при помощи приставки или постфикса глагол можно образовать только от другого глагола, причем дериват останется в том же классе, что и исходный глагол. При помощи суффикса глагол можно образовать от слов любой части речи. Глаголы, образованные суффиксальным способом от других глаголов, переходят в другой класс. Каждый суффикс "приписан" к своему классу, и если образование новых глаголов при помощи этого суффикса продолжается, класс оказывается продуктивным. Класс НУ, например, является продуктивным именно за счет того, что туда попадают новые глаголы, образованные при помощи суффикса *-ну-* и обозначающие однократное действие. Непродуктивные классы представляют собой фиксированное число бесприставочных глаголов и те глаголы, которые образованы от них при помощи

префиксов или постфиксов. Важно также иметь в виду, что у исходного глагола и образованных от него приставочных в абсолютном большинстве случаев совпадают все особенности спряжения. Следовательно, список бесприставочных глаголов позволяет полностью представить структуру глагольных классов и их особенности.

На данном этапе мы составили списки бесприставочных глаголов для всех непродуктивных классов и для класса ЕЙ. При подсчете мы столкнулись с многочисленными проблемами, связанными с морфологическим членением некоторых глаголов, так что для остальных продуктивных классов подсчеты еще не закончены. Список бесприставочных глаголов тех классов, для которых уже завершены подсчеты, находится в Приложении 6. Те глагольные основы, которые представлены в языке только в сочетании с префиксами, обозначаются, например, как *-бегнуть* и условно считаются за один бесприставочный глагол. В этом списке, в частности, можно видеть, что, хотя в А классе больше глаголов, чем в ЕЙ (842 : 638), класс А содержит всего 91 бесприставочный глагол, в то время как класс ЕЙ – 252.

Высчитав объем глагольных классов, мы заинтересовались не только количеством бесприставочных глаголов в каждом классе. Видя, насколько класс АЙ превосходит все прочие глагольные классы, мы захотели понять, за счет чего это получается. Иначе говоря, что за глаголы входят в этот и в другие продуктивные классы? От каких частей речи и при помощи каких формантов они образованы? Обусловлена, ли, скажем, величина класса АЙ тем, что существует много суффиксов, позволяющих образовать глаголы этого класса, или тем, что посредством некоторых суффиксов можно образовать сотни, если не тысячи глаголов? Кроме того, нас интересовало также, много ли в продуктивных классах непродуктивных глаголов.

Эта часть работы еще не доделана. Некоторые сведения по этому поводу содержались в Академической грамматике. Там был указан ряд суффиксов, при помощи которых образуются глаголы продуктивных классов, а также даны списки немотивированных глаголов некоторых продуктивных классов. Вся эта информация содержится в Приложении 6. Однако при последующей работе с базой данных предстоит заполнить пробелы этого списка: составить списки немотивированных глаголов ОВА и И-классов и перечислить все суффиксы, при помощи которых образуются мотивированные глаголы, указав их приблизительное значение.

Так как огромный массив данных пока еще не обработан, мы можем только привести несколько примеров того, что это нам может дать. Наиболее наглядная иллюстрация касается уже обработанного класса АЙ. Почему он такой большой? В него вошло больше трети от общего числа глаголов русского языка. Ни в Академической грамматике, ни в "Грамматическом словаре русского языка" не объясняется, чем именно это обусловлено. Несколько студентов-филологов, опрошенных нами, смогли сказать только, что, видимо, в этот класс просто входит много разнообразных производных и непроизводных глаголов. На самом же деле существует вполне рациональное объяснение. Наиболее распространенный случай суффиксации – образование глагола несовершенного вида от глагола совершенного вида. Все глаголы, образованные таким образом, попадают в АЙ-класс. Это глаголы на *-ать (-ять), -ивать (-ывать) и -вать*.

Исключения (которые мы нашли, пользуясь базой данных):

СВ	Класс	НСВ	Класс
<i>-знать</i>	АЙ	<i>-знавать</i>	АВАЙ
<i>-стать</i>	Н	<i>-ставать</i>	АВАЙ
<i>дать</i>	ANOMAL	<i>давать</i>	АВАЙ
<i>-вести</i>	З-С	<i>-водить</i>	И
<i>-нести</i>	З-С	<i>-носить</i>	И
<i>-везти</i>	Д-Т	<i>-возить</i>	И
<i>-поведать</i> (<i>за-, ис-(-ся), про-</i>)	АЙ	<i>-поведовать</i> (<i>за-, ис-(-ся), про-</i>)	ОВА
<i>стать</i>	Н	<i>становиться</i>	И
<i>идти</i>	IRREG (Д-Т)	<i>ходить</i>	И
<i>-ложить</i>	И	<i>-кладывать</i> (просто необычная видовая пара)	АЙ

Во всех видовых парах глаголы несовершенного вида вторичны. Исключением могли бы быть пары, где глагол совершенного вида образуется при помощи приставки (при условии, что он не приобретает дополнительного значения). Однако Зализняк вообще не учитывает такие пары как видовые. Поэтому в нашей базе данных это утверждение верно для всех видовых пар. В результате огромное количество глаголов совершенного вида, распределенных по всем глагольным классам русского языка, имеет парный глагол несовершенного вида, относящийся к АЙ-классу. Естественно, он пополняется и за счет других производных глаголов (подробнее см. Приложение 6), однако это уже меньше влияет на размер класса.

Полученные нами из базы данных сведения об акцентных парадигмах и о корреляции между ними и глагольными классами мы выделили в отдельную подглаву.

3. Сведения об акцентных парадигмах.

В базе данных мы в основном сохранили обозначения акцентных парадигм, принятые в словаре Зализняка. Они приводятся в Приложении 5. Там же содержатся указания на произведенные нами небольшие изменения и прочие необходимые сведения. В базе данных указания на акцентные парадигмы помещены в столбцах PRAES и PAST (акцентные парадигмы Наст./буд. и Прош. вр. соответственно).

Вопреки распространенному мнению о том, что в русском языке существует очень слабая корреляция между акцентными парадигмами и глагольными классами²⁰, в основном для глаголов одного класса бывает характерно одно единственное сочетание акцентных парадигм Наст./буд. вр. и Прош. вр. Например, все глаголы **О**-класса имеют в Наст./буд. вр. акцентную парадигму **c**, а в Прош. вр. – акцентную парадигму **a**. Однако часто это остается совершенно неочевидным при просмотре таблиц, так как общую картину «портят» различные исключения, лексические (отдельные глаголы и их производные) и грамматические. К последним относятся все глаголы с приставкой *вы-*. Эта приставка перетягивает на себя ударение, поэтому всем глаголам на *вы-* свойственно сочетание акцентных парадигм **a/a**.

Заметим также, что акцентная парадигма прош. вр. **/c'** встречается только у возвратных глаголов. Это не отдельная самостоятельная акцентная парадигма, а своего рода вариант парадигмы **/c** для возвратных глаголов. Если глагол без постфикса *-ся* имел, например, сочетание акцентных парадигм **b/c**, то образованный от него глагол с *-ся* будет иметь **b/c'**.

Здесь следует сказать о преимуществах выбранной нами Одноосновной системы при "вылавливании" из общей массы глаголов с различными аномалиями, в том числе и в акцентных парадигмах. В Одноосновной системе некоторые глаголы хотя и отнесены к одному из 24 классов, выделены из общей массы как "неправильные" (irregular). Кроме того, есть также глаголы "с аномалиями" (anomalous), они же атематические, которые

²⁰ См., например, Зализняк 1985, Редькин 1971

составляют отдельную группу и не входят в 24 регулярных класса. В Академической грамматике и у Зализняка также есть группа неправильных глаголов, но она включает в себя и атематические, и глаголы с теми или иными неправильностями в спряжении. Некоторые глаголы, которые в Одноосновной системе считаются неправильными, распределены по регулярным классам. Таким образом, когда мы пытаемся выявить корреляцию между глагольными классами и акцентными парадигмами, у нас не получается сколько-нибудь удовлетворительного соответствия.

Корреляция между акцентными парадигмами и глагольными классами показана в помещенной ниже таблице. В данном случае нам оказалось удобнее разбирать акцентные парадигмы Наст./буд. и Прош. вр. в паре, а не по отдельности, как они указаны в базе данных. Сокращение "а.п." значит "акцентная парадигма". Кроме того, в таблице введено понятие "спорные глаголы". Так обозначены глаголы, которые у Зализняка и Якобсона относятся к разным классам (это само по себе уже указывает на то, что они не вполне подходят ни для одного, ни для другого класса).

назв.	а.п.	кол-во	комментарии
АЙ	a/a	11735	
ЕИ	a/a	638	
Й	a/a	32	основное сочетание а.п.: <i>почить, дуть, обуть, разуть</i> (и произв.) + <i>выгнить</i>
	b/c	13	<i>гнить</i> (и производные)
ОЙ	a/a	98	
ОВА	a/a	2730	также у глаголов с приставкой <i>вы-</i>
	b/a	85	«короткие» глаголы: <i>жевать, блевать, клевать, плевать, ковать, совать, сновать</i> (и произв.) ²¹
Н	a/a	59	
НУ	a/a	431	Эти 2 сочетания а.п. можно считать основными (сочетание <i>a/a</i> также у глаголов с приставкой <i>вы-</i>).
	b/a	593	
	c/a	48	<i>обмануть, минуть, тонуть, глянуть, помянуть, тянуть</i> (и произв.)
(НУ) ^a	a/a	305	
(НУ) ^b			
Д-Г	b/b	135	
	b/a	36	<i>насть, красть</i> (и производные)

²¹ Это все глаголы ОВА-класса, имеющие односложную основу наст. вр.

З-С	a/a	16	глаголы с приставкой <i>вы-</i>
	b/b	127	
	a/a	30	<i>лезть, отверзть</i> (и производные) + глаголы с приставкой <i>вы-</i>
Б	b/a	23	<i>грызть</i> (и производные)
	b/b	36	
Г-К	a/a	3	глаголы с приставкой <i>вы-</i>
	b/b	214	
	b/a	22	<i>стричься</i> (и производные)
	c/b	10	<i>мочь</i> (и производные)
n/s А	a/a	11	глаголы с приставкой <i>вы-</i>
	b/c	103	
	(b/c'')	73	возвратные
	b/a	21	«спорные» <i>ржать, слать</i> (и производные) + <i>попрать</i> и <i>воззвать</i>
	c/a	16	«спорный» <i>стлать</i> (и производные)
В	a/a	18	глаголы с приставкой <i>вы-</i>
	b/c	34	
	(b/c'')	13	возвратные
(И)Й	a/a	3	глаголы с приставкой <i>вы-</i>
	b/a	62	<i>бить, шить</i> (и производные)
	b/c	54	<i>вить, лить, пить</i> (и производные)
	(b/c'')	39	возвратные
/Р	a/a	8	глаголы с приставкой <i>вы-</i>
	b/a	62	
	b/c	8	<i>мереть</i> (и производные) + -* <i>переть</i> (<i>запереть, опереть, отпереть</i>)
	(b/c'')	1	<i>запереться</i>
	b/b	1	<i>отпереться</i> 'открыться'
/М-Н	a/a	5	глаголы с приставкой <i>вы-</i>
	b/a	78	
	b/c	3	<i>за-чать, на-чать, по-чать</i> (исторически один и тот же корень)
	b/b	2	<i>за-чаться, на-чаться</i>
Й/М	a/a	4	глаголы с приставкой <i>вы-</i>
	b/c	12	
	(b/c'')	1	один возвратный <i>уняться</i> (остальные не по правилам имеют <i>b/b</i>)
НИМ	b/b	4	остальные возвратные (<i>обняться, (при)подняться, разняться</i>)
	c/c	11	
НИМ	(c/c'')	7	возвратные

А	a/a	319	Эти 2 сочетания а.п. можно считать основными (сочетание <i>a/a</i> также у глаголов с приставкой <i>вы-</i>)
	c/a	481	
	b/a	41	смеяться, вопиять + «спорные» сосать, орать (и производные)
И	<i>a/a</i>	2923	Все 3 сочетания а.п. можно считать основными (сочетание <i>a/a</i> также у глаголов с приставкой <i>вы-</i>)
	<i>b/a</i>	2396	
	c/a	1556	
Е	b/a	228	
	a/a	33	<i>обидеть, видеть, зависеть</i> (и произв.) + глаголы с приставкой <i>вы-</i>
	c/a	63	<i>терпеть, вертеть, смотреть</i> (и производные)
ЖА	b/a	191	
	a/a	16	<i>слышать</i> (и производные) + глаголы с приставкой <i>вы-</i>
	c/a	40	<i>держаться, дышать</i> (и производные)
АВАЙ	b/a	93	также глаголы с приставкой <i>вы-</i> (это случай-исключение)
О	c/a	66	
	a/a	5	глаголы с приставкой <i>вы-</i>

Чередования акцентных парадигм:

И	b//c / a	117	
	c//b / a	182	
n/s А	b / c//a	15	<i>ткать</i> и производные
	(b / c''//a)	1	
Д-Т	b / c//a	10	<i>прясть</i> (и невозвратные производные)
	b / b//a	2	возвратные производные
Е	c//b / a	2	<i>призреть, узреть</i>
/Р	b / b//a	1	<i>опереть</i>
	b / c//a	1	<i>перемереть</i>
	(b / c''//a)	1	<i>опереться</i>

Кроме того, у глаголов, для которых возможны формы, образованные по модели двух разных классов, также обычно встречаются чередования акцентных парадигм. Так

как таких глаголов немного и чередуются у них основные акцентные парадигмы соответствующих классов, мы не будем приводить здесь таблицу.

Классы с тремя основными сочетаниями акцентных парадигм:

Получается, что только для одного глагольного класса **И** характерны сразу три сочетания акцентных парадигм. При этом заметим, что различаются только акцентные парадигмы Наст. вр. (*a*, *b*, *c*), а акцентные парадигмы Прош. вр. совпадают (*a*). Знаменательно, что именно на этот класс приходится максимальное количество чередований акцентных парадигм Наст./буд. вр. В других классах, если такие чередования и есть, они ограничиваются одним корнем и производными от него.

Классы с двумя основными сочетаниями акцентных парадигм:

Й-класс: *b/a* и *b/c* (но там, в сущности, всего 5 производных глаголов)

А-класс: *a/a* и *c/a*

НУ-класс: *a/a* и *b/a*

Для остальных глагольных классов характерно **одно основное сочетание акцентных парадигм.**

Класс Й/М – НИМ следует разобрать особо, так как это своего рода «класс-исключение». Он весь состоит из глаголов, исторически имевших один и тот же корень (*за-нять*, *на-нять* и т.д.), однако претерпевших различные изменения в процессе развития языка. В результате в этом классе есть три разных сочетания акцентных парадигм, не считая сочетаний с */c*». Основные из них два (по одному на подкласс): *b/c* и *c/c*.

Теперь скажем пару слов о ценности подобных данных для психолингвистических экспериментов. Когда мы обрабатывали материалы эксперимента на порождение форм от квазиглаголов, описанного в первой главе, мы даже не решились приступить к анализу акцентных парадигм полученных нами форм. Ведь мы не располагали никакой информацией о том, какие акцентные парадигмы "нормальны" для данного класса, а какие характеризуют лишь отдельные глаголы. Анализ экспериментальных данных с учетом

выявленных нами закономерностей, скорее всего, позволил бы в ряде случаев с большей уверенностью говорить об использовании не общего правила, а аналогии с конкретным глаголом. Кроме того, интересно было бы проследить, насколько корреляция между акцентными парадигмами и глагольными классами в ответах испытуемых соответствует тем закономерностям, которые действуют в языке. И наконец, созданная нами матрица может оказаться незаменимой при составлении анкет для эксперимента. Она позволяет понять, какие ударения характерны для глаголов того или иного класса, и, в зависимости от желания, выбрать "типичных представителей" класса или глаголы-исключения.

4. Заключение.

Одним из главных итогов нашей работы представляется создание электронной базы данных на основе "Грамматического словаря русского языка". База данных содержит 27 408 глаголов со всеми их словоизменительными характеристиками. Этот материал рассортирован по столбцам таким образом, чтобы максимально облегчить поиск, сортировку и статистическую обработку. Структура базы данных сильно отличается от исходной структуры словаря, так как многие формы и особенности спряжения были классифицированы заново, более удобными и полезными для дальнейшего их анализа способами.

В результате мы смогли узнать объем русских глагольных классов. При этом был сделан перевод из глагольной классификации Зализняка в классификацию, предложенную Якобсоном, так называемую Одноосновную систему, что дало нам ряд важных преимуществ.

Мы также начали подсчет бесприставочных и непроизводных основ для каждого класса. Эта часть работы на настоящий момент еще не полностью завершена. Вся эта информация ни в каком виде не содержалась в словаре Зализняка и была собрана нами самостоятельно. Подобные сведения показались нам ценными и интересными, так как они дают нам качественно новое представление о том, как устроена система русских глагольных классов. Теперь мы не только располагаем данными относительно их объема и продуктивности, но также можем понять, "что стоит" за этими характеристиками: какие словообразовательные модели делают тот или иной класс продуктивным, за счет каких глаголов некоторые классы значительно превосходят другие по объему.

На материале разработанной нами базы данных мы проследили корреляцию между акцентными парадигмами и глагольными классами, и выявили все случаи, так или иначе не соответствующие общей закономерности. Получить эти данные из словаря Зализняка было невозможно, причем не только из-за колоссальных подсчетов, которые пришлось бы произвести вручную. Система глагольных классов у Зализняка и, правда, в меньшей степени, в Академической грамматике построена таким образом, чтобы максимальное количество глаголов с нестандартными парадигмами попало в регулярные классы, а не рассматривалось как исключения. Из-за этого почти во всех глагольных классах оказалось представлено несколько акцентных парадигм, и создавалось неверное впечатление о том, что в русском языке существует очень слабая корреляция между глагольным классом и типом акцентной парадигмы глагола.

Кроме того, нами были получены статистические данные относительно видовых пар, собрана информация о различных чередованиях в глагольных основах и т.д. Теперь все эти данные легко найти и проанализировать статистически.

Полученные нами сведения уже были применены в рамках нашего проекта для объяснения экспериментального материала и при составлении анкет. База данных будет полезна любому исследователю, изучающему русскую глагольную морфологию.

V. Библиография.

1. Академическая грамматика 1980. "Русская грамматика". М., 1980.
2. Зализняк А.А. 1985. От праславянской акцентуации к русской. М., 1985.
3. Зализняк А.А. 1987. Грамматический словарь русского языка: Словоизменение. Ок. 100 000 слов. М., 1987.
4. Засорина Н.В. 1969. Частотный словарь современного русского языка. М., 1969.
5. Редькин В.А. 1971. Акцентология современного русского языка. М., 1971.
6. Слюсарь Н.А. 2000а. Порождение русских глагольных словоформ носителями русского и английского языков // II Межвузовская научная конференция студентов-филологов. 5-9 апреля 1999 г. Тезисы докладов. Ч.2. СПб., 2000. С.34.
7. Слюсарь Н.А. 2000б. Порождение словоформ от русских квазиглаголов // III межвузовская научная конференция студентов-филологов. 10-14 апреля 2000 г. Тезисы докладов. Ч.2. СПб., 2000. С. 27–28.
8. Слюсарь Н.А. 2001. База данных по русскому глаголу на основе "Грамматического словаря русского языка" // IV Межвузовская научная конференция студентов-филологов. 9-13 апреля 2001 г. Тезисы докладов. СПб., 2001. С.89.
9. Слюсарь Н.А. 2003. Морфологический анализ или извлечение из памяти: эксперимент с русскими глагольными формами // VI Межвузовская научная конференция студентов-филологов. 7-11 апреля 2003 г. Тезисы докладов. СПб., 2003. С.106.
10. Черниговская Т.В., Гор К., Клиорина Т., Слюсарь Н.А. 1999. Правила в речевой деятельности: порождение глагольных форм в родном и неродном языке // Материалы XXVIII межвузовской научно-методической конференции преподавателей и аспирантов. Вып.16. Секция общего языкознания. Ч.2. СПб., 1999. С. 67–71.
11. Alegre, M., and Gordon, P. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40, 41-61.
12. Baayen, R.H., Dijkstra, T., and Schreuder, R. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, 36, 94–117.
13. Bortolini, U., Leonard, L.B., and Caselli, M.C. 1998. Specific Language Impairment in Italian and English: Evaluating Alternative Accounts of Grammatical Deficits. *Language and Cognitive Processes*, 13, 1-20.

14. Bybee, J.L. 1985. *Morphology: A study of the relation between meaning and form*. Philadelphia: John Benjamins Publishers.
15. Bybee, J.L. 1988. Morphology as lexical organization. Hammond, M., and Noonan, M. (eds.). *Theoretical morphology: Approaches in modern linguistics*. San Diego, CA: Academic Press.
16. Bybee, J.L. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10, 425-455.
17. Caramazza, A., Laudanna, A., and Romani, C. 1988. Lexical access and inflectional morphology. *Cognition*, 28, 297–332.
18. Chernigovskaya, T., and Gor, K. 2000. The Complexity of Paradigm and Input Frequencies in Native and Second Language Verbal Processing: Evidence from Russian. *Journal of Verbal Learning and Verbal Behavior*, 3 (2), 20-38.
19. Chernigovskaya, T., and Gor, K. 2002. Generalization of a Symbolic Rule in L2 Verbal Processing. *Formal Description of Slavic Languages (FDSL)*. Leipzig. 26–36.
20. Clahsen, H. 1999. Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22, 991-1060.
21. Clahsen, H., and Almazán, M. 1998. Syntax and morphology in children with Williams Syndrome. *Cognition* 68, 167–98.
22. Clahsen, H., Rothweiler, M., Woest, A., and Marcus, G. 1992. Regular and irregular inflection in the acquisition of German noun plurals. *Cognition* 45, 225–55.
23. Daugherty, K., and Seidenberg, M. 1992. Rules or connections? The past tense revisited. *Proceedings of the 14th annual meeting of the cognitive science society*. Erlbaum, Hillsdale, NJ. 149–156.
24. Ellis, N.C., and Schmidt, R. 1998. Rules or Association in the Acquisition of Morphology? The Frequency by Regularity Interaction in Human and PDP Learning of Morphosyntax. *Language and Cognitive Processes* 13, 307-336.
25. Elman, J.L., Bates, E., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. 1996. *Rethinking Innateness: a Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

26. Embick, D., Hackl, M., Schaeffer, J., Kelepir, M., and Marantz, A.. 2000. A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive Brain Research*, 10, 345-348.
27. Gor, K., and Chernigovskaya, T. 2001. Rules in the Processing of Russian Verbal Morphology. Zybatow, G., Junghanns, U., Mehlhorn, G., and Szucsich, L. (eds.). *Current Issues in Formal Slavic Linguistics*. Frankfurt/Main, . 528-536.
28. Grodzinsky, Y. 1990. *Theoretical perspectives on language deficits*. Cambridge, MA: MIT Press.
29. Hare, M., and Elman, J. 1995. Learning and morphological change. *Cognition* 56, 61–98.
30. Hare, M., Elman, J., and Daugherty, K. 1995. Default generalisation in connectionist networks. *Language and Cognitive Processes* 10, 601–630.
31. Jaeger, J.J., Lockwood, A.H., Kemmerer, D.L., Van Valin, R.D., Murphy, B.W., and Khalak, H.G. 1996. A positron emission tomographic study of regular and irregular verb morphology in English. *Language* 72, 451–97.
32. Jaeger, J.J., Van Valin, R.D. Jr., and Lockwood, A.H. 1998. Response to Seidenberg and Hoeffner. *Language*, 74, 123-128.
33. Langacker, R. 1987. *Foundations of cognitive grammar, vol. 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.
34. Langacker, R. 1988. A usage-based model. Rudzka-Ostyn, B. (ed.), *Topics in cognitive linguistics*. Amsterdam: John Benjamins. 127-161.
35. Lönnngren L. (ed.). 1993. Частотный словарь современного русского языка. (A Frequency Dictionary of Modern Russian. With a Summary in English.). Uppsala, 1993.
36. Lukács, Á., and Pléh, C. 1999. Hungarian cross-modal priming and treatment of nonsense words supports the dual-process hypothesis. *Behavioral and Brain Sciences*, 22, 1030-1031.
37. Lukatela, G., Gligorijević, B., Kostić, A., and Turvey, M.T. 1980. Representation of inflected nouns in the internal lexicon. *Memory and Cognition*, 8, 415-423.
38. MacWhinney, B., and Leinbach, J. 1991. Implementations are not conceptualizations: Revisiting the verb learning model. *Cognition* 40, 121-157.
39. Marchman, V., 1993. Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience* 5, 215–234.

40. Marcus, G.F., Brinkmann, U., Clahsen, H., Weise, R., and Pinker, S. 1995. German inflection: The exception that proves the rule. *Cognitive Psychology* 29, 189-256.
41. Marcus, G.F., Pinker, S., Ullman, M.T., Hollander, M., Rosen, T.J., and Xu, F. 1992. *Overregularization in language acquisition*. Monographs of the Society for Research in Child Development 57 (4), Serial No. 228. Chicago: University of Chicago Press.
42. Marcus, G.F. 1998. Can connectionism save constructivism? *Cognition* 66, 153-182.
43. Marslen-Wilson, W., and Tyler, L. 1997. Dissociating types of mental computation. *Nature* 387, 592-94.
44. Marslen-Wilson, W. 2001. Access to lexical representations. Cross-linguistic issues. *Language and Cognitive Processes* 16, 699-708.
45. McCarthy, J., and Prince, A. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8, 209-283.
46. Orsolini, M., and Marslen-Wilson, W. 1997. Universals in morphological representation: Evidence from Italian. *Language and Cognitive Processes* 12, 1-47.
47. Orsolini, M., Fanari, R., and Bowles, H. 1998. Acquiring regular and irregular inflection in a language with verb classes. *Language and Cognitive Processes* 13, 425-464.
48. Pinker, S. 1991. Rules of Language. *Science* 253, 530-535.
49. Pinker, S. 1995. Why the child holded the baby rabbits: A case study in language acquisition. Gleitman, L.R., and Liberman, M. (eds.). *An Invitation to Cognitive Science. Vol. 1. Language*. Cambridge, MA: MIT Press.
50. Pinker, S. 1999. *Words and Rules: the Ingredients of Language*. New York, NY: Basic Books.
51. Pinker, S., and Prince, A. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28, 73-193.
52. Pinker, S., and Prince, A. 1991. Regular and irregular morphology and the psychological status of rules of grammar. *Proceedings of the 1991 Meeting of the Berkeley Linguistics Society*.
53. Pinker, S., and Prince, A. 1994. Regular and irregular morphology and the psychological status of rules of grammar. Lima, S.D., Corrigan, R.L., and Iverson, G.K. (eds.). *The reality of linguistic rules*. Amsterdam: John Benjamins. 353-388.

54. Plunkett, K. 1998. Language Acquisition and Connectionism. *Language and Cognitive Processes* 13, 97-104.
55. Plunkett, K., and Marchman, V. 1991. U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition* 38, 43-102.
56. Plunkett, K., and Marchman, V. 1993. From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition* 48, 21-69.
57. Prasada, S., and Pinker, S. 1993. Generalization of regular and irregular morphological patterns. *Language and cognitive processes* 8, 1-56.
58. Ragnarsdóttir, H., Simonsen, H.G., and Plunkett, K. 1999. The acquisition of past tense morphology in Icelandic and Norwegian children: an experimental study. *Journal of Child Language* 26, 577-618
59. Rumelhart, D.E., and McClelland, J.L. 1986. On learning the past tenses of English verbs. McClelland, J.L., and Rumelhart, D.E. (eds.). *Parallel distributed processing: Explorations in the microstructures of cognition*, vol. 2. Cambridge, MA: Bradford / MIT Press. 216-271
60. Seidenberg, M.S., and Hoeffner, J.H. 1998. Evaluating behavioral and neuroimaging data on past tense processing. *Language* 74, 104-122.
61. Simonsen, H.G., and Bjerkan, K.V. 1998. Testing past tense inflection in Norwegian: A diagnostic tool for identifying SLI children? *International Journal of Applied Linguistics* 8 (2), 251-270.
62. Slioussar, N. 2001. Mental Lexicon and Russian Verb Morphology: Experimental Study. Niemi, J., and Heikkinen, J. (eds.). *Nordic and Baltic Morphology. Papers from a NorFA Course, Tartu, June 2000*. Studies in Languages 36, Faculty of Humanities, University of Joensuu.
63. Stemberger, J., and MacWhinney, B. 1988. Are inflected forms stored in the lexicon? Hammond, M., and Noonan, M. (eds.). *Theoretical morphology: Approaches in modern linguistics*. New York: Academic Press. 101-116
64. Taft, M. 1979. Recognition of affixed words and the word frequency effect. *Memory and Cognition* 7, 263-272.

65. Ullman, M.T. 1999. Acceptability ratings of regular and irregular past-tense forms: evidence for a dual-system model of language from word frequency and phonological neighborhood effects. *Language and Cognitive Processes* 14, 47-67.
66. Ullman, M.T., Bergida, R., and O'Craven, K. 1997a. Distinct fMRI activation patterns for regular and irregular past tense. *NeuroImage* 5, 555.
67. Ullman, M.T., Corkin, S., Coppola, M., Hickock, G., Growdon, J.H., Koroshetz, W.J., and Pinker, S. 1997b. A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience* 9, 266-76.
68. Ullman, M.T., and Gopnik, M. 1999. Inflectional morphology in a family with inherited specific language impairment. *Applied Psycholinguistics* 20, 51-117.
69. Van der Lely, H.K.J., and Ullman, M.T. 1996. The Computation and Representation of Past Tense Morphology in Specifically Language Impaired and Normally Developing Children. *BUCLD 20 Proceedings*. 804-815.

VI. Приложения.

Приложение 1. Одноосновная Система (OSM).

One Stem Morphology (OSM), или Одноосновная Система, была разработана Р.О.Якобсоном и его последователями. Она представляет собой один из вариантов описания глагольной системы русского языка, альтернативный по отношению к традиционному подходу, который представлен в Академической Грамматике (АГ). В отличие от АГ, в OSM, как это можно вывести из самого ее названия, выделяется всего одна глагольная основа, которая называется basic stem (BS).

Определить BS у каждого конкретного глагола можно следующим образом. Это более длинная из двух традиционно выделяемых глагольных основ – основы инфинитива и основы наст. вр., определяемой по форме 3 л. мн. ч. Если длина этих двух основ совпадает, выбирается основа настоящего/будущего времени. Глаголы делятся на классы в зависимости от типа BS. Для каждого глагольного класса выведены правила, согласно которым от BS можно образовать инфинитив и формы основа настоящего/будущего времени. Основной принцип действия этих правил заключается в том, что, когда к BS прибавляются суффиксы и окончания, происходит сокращение сочетаний из двух гласных или двух согласных. Таким образом, если BS оканчивается на гласный, он остается в инфинитиве, но выпадает в основе настоящем/будущем времени. Если BS оканчивается на согласный, то в настоящем/будущем времени окончания прибавляются прямо к BS, а в инфинитиве этот согласный отсутствует.

Выделяется 23 глагольных класса, а также Irregular Verbs (неправильные глаголы) и Anomalous Verbs (в эту группу включены глаголы *есть*, *надоесть*, *дать*, *создать* и другие глаголы с компонентами *-есть* и *-дать*). Всем глагольным классам присвоены условные названия. Ниже дается классификация BS и названия классов, в которые входят глаголы с BS того или иного типа. Далее следует таблица, в которой каждый глагольный класс иллюстрируется одним примером.

Basic Stems	
Vocalic (оканчиваются на гласный) Совпадают с основой инфинитива	Consonantal (оканчиваются на согласный) Совпадают с основой наст. вр.

Suffixed (с суффиксом)		Suffixed	Non-suffixed (без суффикса)		
Глаголы 1 спр.	Глаголы 2 спр. ²²		Resonant (основы на сонанты и -e-)		Obstruent (основы на шумные)
			Syllabic (слоговые) ²³	Non-syllabic (неслоговые)	
A; n/s A; ОВА; О; НУ; (НУ)	И; Е; ЖА	АЙ; ЕЙ	В; Н; Й; ОЙ; ИЙ	/Р; /М-/Н; Й/М; НИМ	Д-Т; З-С; Г- К

Exceptional type: АВАЙ.

№	название	пример
1	В	жи-л-и – жив-ут, жив-у
2	Н	де-л-и – ден-ут, ден-у
3	Й*а	ду-л-и – дуј-ут, дуј-у
4	ОЙ	мы-л-и – мој-ут, мој-у
5	ИЙ*β	пи-л-и – пј-ут, пј-у
6	/Р	тер-л-и – тр-у, тр-ут
7	/М-/Н	жа-л-и – жм-у, жм-ут
8	Й/М*γ НИМ	поня-л-и – пойм-ут, пойм-у сня-л-и – сним-ут, сним-у
9	Д-Т	ве-л-и – вед-ут, вед-у
10	З-С	вез-л-и – вез-ут, вез-у
11	Б	греб-л-и – греб-ут, греб-у
12	Г-К	пек-л-и – пек-ут, пек-у
13	И	проси-л-и – прос'-ат, прош-у
14	Е	виде-л-и – вид'-ат, виж-у
15	ЖА²⁴	держа-л-и – держ-ат, держ-у
16	А	писа-л-и – пиш-ут, пиш-у
17	n/s А²⁵	жда-л-и – жд-ут, жд-у

²² Все остальные глагольные классы, с BS как на гласный, так и на согласный, включают глаголы 1 спряжения.

²³ Под "слоговыми" и "несловowymi" основами подразумевается следующее. Все приставочные глаголы, входящие в эти классы, либо имеют односложную основу инфинитива, либо ясно видно, как, например, в случае с глаголом начать, что они имели ее исторически. В наст./буд. вр. часть этих глаголов имеет односложную основу, и у них выделяется слоговая BS, а у других основа представляет собой сочетание согласных, как, например, у глагола тереть (трут) или мять (мнут), и их BS называется неслоговой.

²⁴ Буква Ж в данном случае символизирует звуки |ж|, |ш|, |ч|, |ш':| и |j|.

18	ОВА	требова-л-и – требуј-ут, требуј-у
19	О	коло-л-и – кол'-ут, кол'-у
20	НУ	толкну-л-и – толкн-ут, толкн-у
21	(НУ)^{a 26}	слеп-л-и – слепн-ут, слепн-у
	(НУ)^b	сверг-л-и – свергн-ут, свергн-у
22	АЙ	дела-л-и – делај-ут, делај-у
23	ЕЙ	уме-л-и – умеј-ут, умеј-у
24	АВАЙ	дава-л-и – дај-ут, дај-у

Также выделяются **Irregular Stems** (глаголы с определенными аномалиями в образовании форм, которые, однако, все же можно отнести к одному из вышеперечисленных 23 классов) и **Anomalous Stems** (глаголы, которые не входят в систему классов и стоят особняком).

Irregular Stems:

Н	<i>клясть</i>
Й	<i>брить</i>
Й	<i>петь</i>
Й/М – НИМ	<i>взять</i>
Й/М – НИМ	<i>принять</i>
Д-Т	<i>быть</i>
Д-Т	<i>ехать</i>
Д-Т	<i>идти</i>
Д-Т	<i>сесть</i>
Д-Т	<i>расти</i>
Г-К	<i>лечь</i>
Г-К	<i>толочь</i>
Б	<i>ошибиться</i>
Б	<i>ушибить</i>
И	<i>родить</i>
И	<i>родиться</i>
И	<i>читать</i>
Е	<i>реветь</i>
Е	<i>хотеть</i>
ЖА	<i>бежать</i>
n/s А	<i>гнать</i>
n/s А	<i>спать</i>

²⁵ Этот класс обозначен буквосочетанием n/s А, так как туда входят глаголы на *-ать* либо с неслоговой основой наст. вр. (см. сноску к термину Non-syllabic), либо с односложной, в которой появляется беглый гласный.

²⁶ Это НУ чаще выпадает. Сравните "ослепший" и "свергнувший".

О	<i>МОЛОТЪ</i>
----------	---------------

Anomalous Stems:

есть, надоестъ, дать, создать и однокоренные

К характеристикам каждого отдельного класса также относятся

- Наличие регулярных чередований согласных и гласных в глагольных корнях.
- Наличие регулярных суффиксальных чередований. Регулярное выпадение того или иного суффикса в определенных формах.
- Имеют ли глагольные основы данного класса stem stress. Stem stress – ударение, стоящее на основе и не перемещающееся на окончания.

**Приложение 2. Соотношение Академической Грамматики и
Одноосновной Системы (OSM).**

Академическая Грамматика	OSM
I-й класс: соотношение <i>нуль согласной</i> – j	
1-й подкласс: соотношение a – aj <i>играть: игра-л-а – играј-ут</i>	АЙ
2-й подкласс: соотношение e – ej <i>белеть: беле-л-а – белеј-ут</i>	ЕЙ
3-й подкласс: соотношение u – uj <i>почить: почи-л-а – почиј-ут</i>	Й
4-й подкласс: соотношение y – yj <i>дуть: ду-л-а – дуј-ут</i>	Й
5-й подкласс: различие конечн. гл. в основах наст. вр. и прош. вр.	
А подкласс: соотношение u – j <i>бить: би-л-а – бј-ут</i>	ИЙ
Б подкласс: соотношение u – ej <i>брить: бри-л-а – бреј-ут (I)</i>	Игг. Й
В подкласс: соотношение ы – oj <i>выть: вы-л-а – вој-ут</i>	ОЙ
Г подкласс: соотношение e – oj <i>петь: пе-л-а – пој-ут (I)</i>	Игг. Й
Д подкласс: соотношение a – uj <i>живописать: живописа-л-а – живописуј-ут (I)</i>	нет такого
II-й класс: соотношение <i>ова</i> – uj <i>командовать: командова-л-а – командуј-ут</i>	ОВА
III-й класс: соотношение ну – н <i>двинуть: двину-л-а – двин-ут</i>	НУ
IV-й класс: соотношение <i>нуль</i> – н <i>глохнуть: глох-л-а – глохн-ут</i>	(НУ) ^a (НУ) ^b
V-й класс: соотношение <i>гласная</i> – <i>нуль</i>	
1-й подкласс: соотношение a – <i>нуль</i>	
А подкласс: чередование последних согласных основы прош. и наст. вр. <i>махать: маха-л-а – маш-ут</i>	А
Б подкласс: основа наст. вр. оканчивается на j (соотношение ja – j) <i>смеяться: смеја-л-а-ась – смеј-ут-ся</i>	А
В подкласс: основа наст. вр. на ТВ. СОГЛ.	n/s A ²⁷

²⁷ Глаголы *жаждать, орать* "кричать", *сосать, стонать*, относящиеся в АГ к классу V.1.B, не входят в класс n/s A, они относятся к классу А. Зато в класс n/s A входят глаголы *слать*

и чередования гласных нет врать: вра-л-а – вр-у	
Г подкласс: глаголы с беглыми гласными в основе наст. вр. брать: бра-л-а – бер-ут	n/s A
2-й подкласс: соотношение o – нуль полоть: поло-л-а – пол'-ут	O
3-й подкласс: соотношение e – нуль реветь: реве-л-а – рев-ут (I)	Irr. E
VI-й класс: основы наст. и прош. вр. совпадают.	
1-й подкласс: основы с конеч. согл. к – ч , г – ж влечь: влеку – влечешь, стеречь: стерегу – стережешь	Г-К
2-й подкласс: основы на согласные б , с , з и р	
А подкласс: инф. на -сти (-сть), -зи (-зь) грести: гробут – гробешь, нести: несут – несеешь, везти: везут – везешь	З-С, Б
Б подкласс: инф. на -еть тереть: трют – трешь	/Р
В подкласс: инфинитив оканчивается на -ить -шибить (за-, от-, под-, пере-, у- и все возвр. глаг. с этим корнем) (I)	Irr. Б
VII-й класс: соотношения нуль – ð , нуль – т , нуль – в , нуль – н 	
1-й подкласс: соотношение нуль – ð , нуль – т вести: ве-л-а – вед-ут, мести: ме-л-а – мет-ут глагол расти ²⁸	Д-Т
2-й подкласс: соотношение нуль – в жить: жи-л-а – жив-ут	В
3-й подкласс: соотношение нуль – н деть: де-л-а – ден-ут	Н
VIII-й класс: соотношение ва – j 	
давать: дава-л-а – даj-у	АВАЙ
IX-й класс: соотношение гласная а – носовая согласная м или н с предшествующим нулем гласной или гласной и 	
1-й подкласс: в основе наст. вр. н жать: жа-л-а – жн-ут	/М – /Н
2-й подкласс: в основе наст. вр. м жать: жа-л-а – жм-ут	/М – /Н
Тут некоторая сумятица с разделением на типы подкласса, так что	Й/М – НИМ

и *стать*, которые в АГ относятся к V.1.A классу, так как у них чередуется твердый и мягкий |л|.

Появление беглого |e| и чередование |с//ш| описываются как нерегулярные.

²⁸ В OSM глагол *расти* считается Irr. Д-Т.

скажем просто, что, помимо <i>жать</i> (/М – /Н), сюда входят глаголы <i>взять, донять, занять, нанять, обнять, перенять, понять, пронять, снять, унять + взять, принять</i> (они Iгг. Й/М – НИМ) ²⁹	
X-й класс: соотношения и – нуль, е – нуль, а – нуль	
1-й подкласс: соотношение и – нуль <i>мылить: мыли-л-а – мыл-ят</i>	И
2-й подкласс: соотношение е – нуль <i>бдеть: бде-л-а – бд-ят</i>	Е
3-й подкласс: соотношение а – нуль <i>ворчать: ворча-л-а – ворч-ат</i> глаголы <i>знать</i> и <i>спать</i> ³⁰	ЖА

АГ:

Изолированные глаголы

- 1) *зиздиться, зыбить(ся), мяукать* (нет в OSM)
- 2) *хотеть, бежать, чтить* (Irregular: Е, ЖА, И-классы соответственно)
- 3) *есть, надоесть, дать, создать* (Anomalous)
- 4) *быть* (а также *забыть, добыть, сбыть* и другие глаголы с компонентом *-быть*), *ехать, идти* (Irregular: Д-Т класс).

OSM:

Irregular stems (глаголы с различными аномалиями, которые все же можно отнести к одному из 24 классов) – смотри список в Приложении 1.

Anomalous stems (*есть, надоесть, дать, создать*).

²⁹ Также сюда относятся всякие *изъять, подъять*...

³⁰ В OSM глаголы *знать* и *спать* считаются Iгг. n/s А.

**Приложение 3. Соотношение системы глагольных классов,
принятых в словаре Зализняка, и Одноосновной Системы (OSM).**

Зализняк	OSM	Как разделять те классы, которые у Зализняка объединены	Что у Зализняка в данный класс не входит, а в OSM входит	Что у Зализняка в данный класс входит, а в OSM не входит
1	АЙ	Одни на <i>-ать(ся)</i> , <i>-ять(ся)</i>		
	ЕЙ	другие на <i>-еть(ся)</i>	<i>греть</i> (З. – кл. 12)	
2	ОВА			
3	НУ			
3⁰	(НУ)^a (НУ)^b	В базе эти подклассы не разведены (но у них разные цифровые пометы, см. Приложение 5)		
4	И			<i>родить(ся)</i> (OSM – Iгг.), <i>читать</i> (OSM – Iгг.)
5	ЖА	Одни на <i>-ать(ся)</i> , <i>-ять(ся)</i> ,		<i>знать</i> (OSM – Iгг.),
	Е	другие на <i>-еть(ся)</i>		<i>спать</i> (OSM – Iгг.), <i>хотеть</i> (OSM – Iгг.), <i>бежать</i> (OSM – Iгг.)
6	А		жаждать, орать ("кричать"), сосать, стонать (З. – кл. 6 ⁰)	<i>ржать,</i> <i>слать,</i> <i>стлать</i> (OSM – n/s A)
6⁰	n/s A		<i>ржать,</i> <i>слать,</i> <i>стлать</i> (З. – кл. 6)	<i>жаждать,</i> <i>орать</i> ("кричать"), <i>сосать,</i> <i>стонать</i> (OSM – A)
7	Д-Т	Все эти глаголы снабжены пометами типа <i>(-т-)</i> , <i>(-с-)</i> и т. п.		<i>сесть</i> (OSM – Iгг.),
	З-С			<i>расти</i> (OSM – Iгг.)
	Б			
8	Г-К			<i>лечь</i> (OSM – Iгг.),

				<i>толочь</i> (OSM – Iгг.)
9	/P			
10	О			<i>молоть</i> (OSM – Iгг.)
11	(И)Й			
12	Й	Одни на -ить(ся) и -уть(ся).		<i>греть</i> (OSM – ЕЙ),
	ОЙ	другие на -ить(ся)		<i>петь</i> (OSM – Iгг.), <i>брить</i> (OSM – Iгг.)
13	АВАЙ			
14	/М-/Н	Надо смотреть каждое слово отдельно (они все разобраны в базе и приведены в Приложении б)		
	Й/М			<i>взять</i> (OSM – Iгг.),
	НИМ			<i>принять</i> (OSM – Iгг.)
15	Н			
16	В			

Словарь Зализняка:

Кроме этих классов, у Зализняка есть **неправильные глаголы**, которые обозначаются пометой Δ:

- 1) *зиждиться, зыбить(ся), густи* (OSM: их нет)
- 2) *есть, надоесть, дать, создать* (OSM: Anomalous)
- 3) *быть, ехать, идти* (OSM: Irregular, Д-Т класс).
- 4) *клясть* (OSM: Irregular, Н класс).
- 5) *–шибить(ся)* (OSM: Irregular, Б класс).
- 6) *реветь* (OSM: Irregular, Е класс).

Глагол *живописать*, для которого в Академической Грамматике выделен отдельный подкласс (I.5.Д) и который отсутствует в OSM, у Зализняка относится к 1 классу.

One Stem Morphology:

Irregular stems (глаголы с различными аномалиями, которые все же можно отнести к одному из 24 классов):

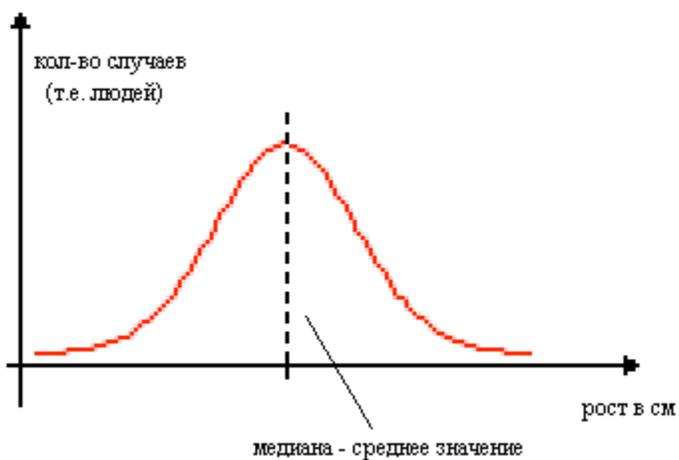
Н	<i>клясть</i> (ЗАЛИЗНЯК – Δ)
Й	<i>брить</i> (ЗАЛИЗНЯК – 12 кл.), <i>петь</i> (ЗАЛИЗНЯК – 12 кл.)
Й/М – НИМ	<i>взять</i> (ЗАЛИЗНЯК – 14 кл.), <i>принять</i> (ЗАЛИЗНЯК – 14 кл.)
Д-Т	<i>быть</i> (ЗАЛИЗНЯК – Δ), <i>ехать</i> (ЗАЛИЗНЯК – Δ), <i>идти</i> (ЗАЛИЗНЯК – Δ), <i>сесть</i> (ЗАЛИЗНЯК – 7 кл.), <i>расти</i> (ЗАЛИЗНЯК – 7 кл.)
Г-К	<i>лечь</i> (ЗАЛИЗНЯК – 8 кл.), <i>толочь</i> (ЗАЛИЗНЯК – 8 кл.)
Б	<i>-шибить(ся)</i> (ЗАЛИЗНЯК – Δ)
И	<i>родить(ся)</i> (ЗАЛИЗНЯК – 4 кл.), <i>читать</i> (ЗАЛИЗНЯК – 4 кл.)
Е	<i>реветь</i> (ЗАЛИЗНЯК – Δ), <i>хотеть</i> (ЗАЛИЗНЯК – 5 кл.)
ЖА	<i>бежать</i> (ЗАЛИЗНЯК – 5 кл.)
n/s А	<i>гнать</i> (ЗАЛИЗНЯК – 5 кл.), <i>спать</i> (ЗАЛИЗНЯК – 5 кл.)
О	<i>молоть</i> (ЗАЛИЗНЯК – 10 кл.)
Й/М – НИМ	<i>взять</i> (ЗАЛИЗНЯК – 14 кл.)
Й/М – НИМ	<i>принять</i> (ЗАЛИЗНЯК – 14 кл.)

Anomalous stems (*есть, надоесть, дать, создать*).

Приложение 4. Статистические методы обработки данных.

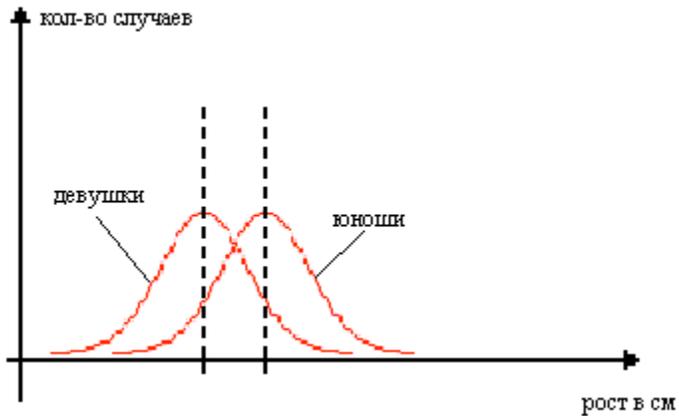
Дисперсионный анализ.

Многие статистические методики, и дисперсионный анализ не исключение, основываются на следующем фундаментальном факте. Известно (и доказано), что множество различных естественных объектов или явлений, характеризующихся каким-то континуальным количественным признаком, представляет собой так называемое нормальное распределение. Возьмем, например, множество всех живущих в данный момент на Земле людей 25 лет и такую их количественную характеристику, как рост (длину человека в сантиметрах). Измерив большую часть этих людей, мы сможем построить график, который будет иметь вид нормального распределения:



Сущность нормального распределения заключается в том, что большая часть объектов группируется вокруг среднего значения (это медиана графика), а по мере отдаления от него количество объектов убывает. Даже интуитивно такая картина представляется вполне естественной. Заметим, что график не имеет соприкосновений с осью абсцисс, иначе говоря, на нем не указано таких значений роста, начиная с которых количество людей будет равно нулю. Это связано с тем, что мы всегда имеем дело только с частью множества. Например, в нашей выборке самым высоким оказался молодой человек ростом 215 см. Но мы не можем ручаться, что в природе не существует другого, несколько более высокого. При этом мы точно знаем, что людей ростом 240 см уже не бывает. Поэтому мы отмечаем на графике ту область, где он начинает стремиться к нулю, но не указываем точных цифр.

Продолжим рассмотрение нашего множества. Во время измерений мы заметили, что девушки оказались в среднем ниже юношей. На графике это можно представить так:



Сущность понятия "в среднем ниже" заключается в том, что значения роста девушек, во многом пересекаясь со значениями роста юношей, группируются вокруг другого, меньшего среднего значения. Иначе говоря, тот факт, что пол является фактором, влияющим на рост, выражается в том, что если разделить людей по половому признаку, средние значения роста (медианы) этих двух групп не совпадут (между ними будет статистически значимая разница). Если же мы разделим наше множество по какому-то признаку, который не является значимым фактором с точки зрения роста, между медианами получившихся групп не будет статистически значимого различия (то есть они почти совпадут или просто будут равны).

Основной целью дисперсионного анализа, которому посвящено это приложение, как раз и является исследование значимости различия между медианами. Может показаться странным, что процедура сравнения средних называется дисперсионным анализом. Это связано с тем, что при исследовании статистической значимости различия между средними нескольких групп мы на самом деле сравниваем дисперсии этих выборок. Дисперсия — это, условно говоря, степень рассеяния элементов, степень их удаленности от медианы.

Для выборки объема n дисперсия вычисляется как сумма квадратов отклонений всех элементов от выборочного среднего, деленная на $n-1$ (объем выборки минус единица). Таким образом, при фиксированном объеме выборки n дисперсия является функцией суммы квадратов отклонений, обозначаемой SS (от *sum of squares*) — чем больше

SS, тем больше дисперсия. В основе дисперсионного анализа лежит разделение дисперсии на части или компоненты. Рассмотрим следующий набор данных:

	Группа 1	Группа 2
Наблюдение 1	2	6
Наблюдение 2	3	7
Наблюдение 3	1	5
Среднее	2	6
SS	2	2
Общее среднее	4	
Общая SS	28	

Средние двух групп существенно различны (2 и 6 соответственно). Сумма квадратов отклонений *внутри* каждой группы равна 2^{31} . Складывая их, получаем 4. Если теперь повторить эти вычисления *без учета* групповой принадлежности, то есть если вычислить SS исходя из общего среднего этих двух выборок, получим величину 28^{32} . Мы видим, что SS (а значит, и дисперсия), основанная на внутригрупповой изменчивости, намного меньше, чем при вычислении на основе общей изменчивости (относительно общего среднего). Причина этого, очевидно, заключается в существенной разнице между средними значениями (она и объясняет существующее различие между суммами квадратов).

Для статистической обработки данных нашего эксперимента мы использовали программу SPSS. Дисперсионный анализ (и его результат – т.н. таблица дисперсионного анализа) в англоязычных статистических пакетах сокращенно называется ANOVA (от *Analysis of Variance*). Это сокращение постепенно заменяет и длинный русскоязычный термин. Чтобы получить таблицу дисперсионного анализа для условного набора данных, который мы используем в качестве примера, надо перегрупповать их следующим образом:

Nabl	Group
2	1
3	1

³¹ SS для группы 1: $2-2=0$, $1-2=-1$, $3-2=1$; $0^2=0$, $(-1)^2=1$, $1^2=1$; $0+1+1=2$.

SS для группы 2: $6-6=0$, $5-6=-1$, $7-6=1$; $0^2=0$, $(-1)^2=1$, $1^2=1$; $0+1+1=2$.

³² SS без учета групповой принадлежности: $1-4=-3$, $2-4=-2$, $3-4=-1$, $5-4=1$, $6-4=2$, $7-4=3$;

$(-3)^2=9$, $(-2)^2=4$, $(-1)^2=1$, $(1)^2=1$, $2^2=4$, $3^2=9$; $9+4+1+1+4+9=28$

1	1
6	2
7	2
5	2

Важно иметь в виду, что зависимые величины и факторы в дисперсионном анализе могут быть только количественными (а не качественными). Если у нас есть всего две группы, мы всегда можем присвоить им условные количественные значения, например, 1 и 2. Если же мы хотим изучить влияние частотности или формы глагола на время реакции, мы можем, скажем, обозначить 2 л. Ед. ч. числом 2, 3 л. Ед. ч. – числом 3, а инфинитив – числом 1, частотность квазислов как 0, редких как 1, а частотных как 2. Итак, ANOVA для нашей группы наблюдений, рассчитанная программой SPSS, будет выглядеть следующим образом:

		Sum of Squares	df	Mean Square	F	Sig.	
Nabl	Main Effects	Group	24	1	24	24	0.008
	Model		24	1	24	24	0.008
	Residual		4	4	1		
	Total		28	5	5.6		

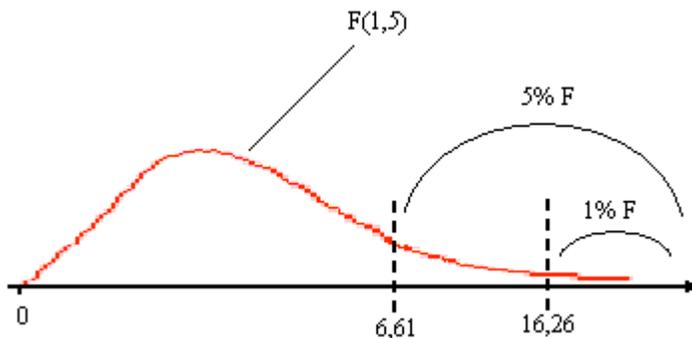
Как видно из таблицы, общая сумма квадратов отклонений $SS = 28$ разбита на компоненты: сумму квадратов, обусловленную внутригрупповой изменчивостью ($2+2=4$; в графе *Residual*) и сумму квадратов, обусловленную различием средних значений между группами ($28-(2+2)=24$; в графе *Main Effects: Group* и в графе *Model*). Внутригрупповая изменчивость (SS) обычно называется остаточной компонентой или дисперсией ошибки (поэтому графа и названа *Residual*, то есть "остаточная"). Это означает, что обычно при проведении эксперимента она не может быть предсказана или объяснена. С другой стороны, так называемую SS эффекта (или компоненту дисперсии между группами) можно объяснить различием между средними значениями в группах. В этой таблице графы *Main Effects* и *Model* совпадают, так как у нас есть всего один фактор. При наличии нескольких факторов для каждого из них будет рассчитана SS эффекта (все они появятся в виде отдельных граф внутри графы *Main Effects*), а их сумма появится в графе *Model*. Это

будет та часть дисперсии, которая объясняется в рамках предложенной нами факторной модели.

Mean Square (MS) – это средний квадрат, то есть *SS*, деленная на число степеней свободы. Число степеней свободы показано в графе *df (Degrees of Freedom)*. В графе *Total* число степеней свободы – это $n-1$, то есть объем выборки минус единица. В графе *Main Effects* это количество значений, которые может принимать данный фактор минус единица. У нас есть две группы, то есть фактор *Group* принимает два значения. Поэтому $df=2-1=1$. *df* в графе *Residual* – это разница между *df* в графе *Total* и в графе *Main Effects*.

Проверка значимости в дисперсионном анализе основана на сравнении компоненты дисперсии, обусловленной межгрупповым разбросом (называемой средним квадратом эффекта или MS эффекта) и компоненты дисперсии, обусловленной внутригрупповым разбросом (называемой средним квадратом ошибки или MS ошибки). Если верна нулевая гипотеза (равенство средних в двух популяциях), то можно ожидать сравнительно небольшое различие выборочных средних из-за чисто случайной изменчивости. Поэтому, при нулевой гипотезе, внутригрупповая дисперсия будет практически совпадать с общей дисперсией, подсчитанной без учета групповой принадлежности. Полученные внутригрупповые дисперсии можно сравнить с помощью F-критерия, проверяющего, действительно ли отношение дисперсий значимо больше 1.

$F = SS \text{ эффекта} / SS \text{ ошибки}$. В нашем случае $F(1,5) = 24 / 1 = 24$. Числа в скобках – это *df*, степени свободы. F – величина, зависящая от степеней свободы. Так, для каждой пары степеней свободы существует график, называемый F-распределением.



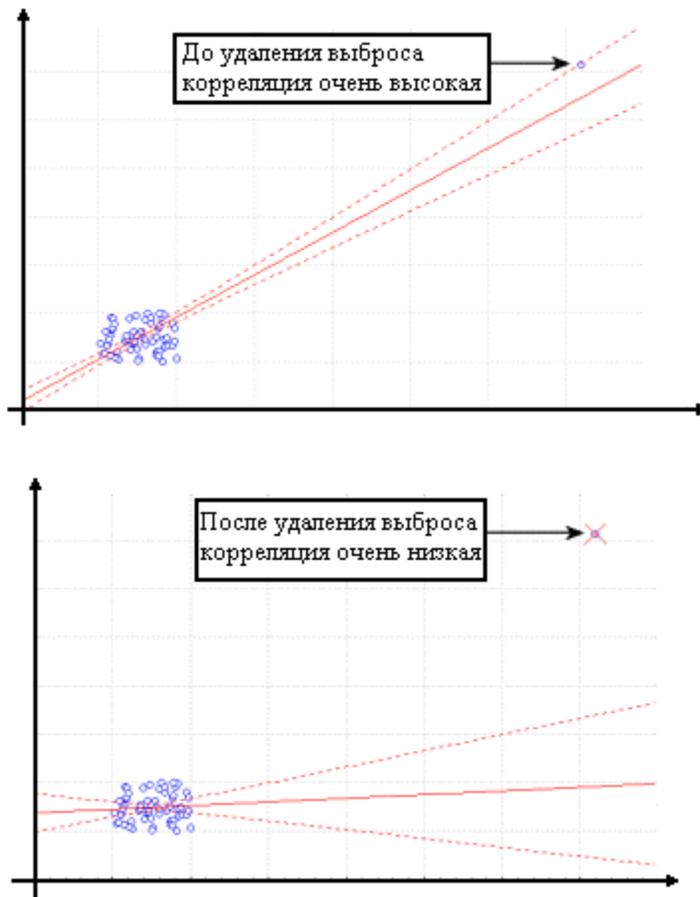
Этот график – множество всех значений F, таких, что различие между внутригрупповой дисперсией и общей дисперсией НЕ является значимым (представление о том, как именно строятся подобные графики, выходит за рамки прикладной статистики).

Предположим, что мы доказываем влияние какого-то фактора на зависимую величину и получили $F(1,5) = 6,61$. Мы знаем, что только 5% F, таких, что различие между внутригрупповой дисперсией и общей дисперсией не является статистически значимым, больше 6,61 (эти числа приводятся в специальных таблицах или рассчитываются при помощи программ типа SPSS). Значит, имея $F(1,5) = 6,61$, мы можем ошибиться с вероятностью 5% (1 к 20). Для того чтобы описывать вероятность ошибки, была введена специальная величина – уровень значимости (p). В SPSS она зачем-то обозначается не p , *Sig.* (от *Significance*). Скажем, если мы можем ошибиться с вероятностью 5%, $p = 0,05$. В рассмотренном выше примере с набором из шести наблюдений F-критерий показывает, что различие между средними значимо на уровне 0,008.

Заметим, что в этом простом примере мы могли бы сразу, причем вручную, вычислить *t-критерий* (критерий Стьюдента) для независимых выборок. Полученные результаты, естественно, совпадут с результатами дисперсионного анализа. Однако дисперсионный анализ содержит гораздо более гибкие и мощные технические средства, позволяющие исследовать планы практически неограниченной сложности. При помощи дисперсионного анализа можно исследовать влияние нескольких взаимодействующих или, наоборот, противодействующих факторов на зависимую величину. Поэтому эта методика и используется в подавляющем большинстве работ, использующих время реакции в качестве зависимой переменной.

Экстремальные значения.

Экстремальные значения, или выбросы, – это нетипичные или редкие значения, которые существенно отклоняются от распределения остальных выборочных данных. Выбросы могут отражать истинные свойства изучаемого явления, а могут быть связаны с ошибками измерения или аномальными явлениями, и поэтому не должны включаться в модель. Ведь всего один выброс может полностью изменить вид зависимости между переменными. Вот яркий пример из учебника статистики. Одна точка выброса обуславливает высокое значение коэффициента корреляции, в то время как на самом деле (в отсутствие выброса) она практически равна нулю:



Для того чтобы найти и удалить выбросы из наших экспериментальных данных, мы использовали специальную функцию в программе SPSS. Если представить действия программы очень упрощенно, можно сказать, что она делает это следующим образом. Сначала рассматривается все множество ответов одного испытуемого, вычисляется среднее значение и общая степень разбросанности (или сгущенности) элементов вокруг среднего. Исходя из этого строится график нормального распределения, более "приплюснутый" или более "вытянутый". Какие-то элементы оказываются слишком далеко от медианы. Исходя из того, что известно о нормальных распределениях, можно сказать, являются ли они аномальными для данного множества. Заметим, что расстройство от среднего значения, начиная с которого элемент может считаться аномальным, зависит от степени разбросанности элементов.

Приложение 5. Значения условных помет и список акцентных парадигм в базе данных.

Значения условных обозначений и помет.

Помета * → * (в столбце)*

У глаголов с такой пометой происходит чередование беглой гласной *o* с нулем на конце приставки. Это свойство определенных глагольных корней. Эта помета встречается только в тех глагольных классах, где есть глаголы с неслоговой основой: **n/s A, /R, (I)J, /M-/N, J/M – NIM**. Кроме того, в классе **G-K** есть один глагол с неслоговой основой – *жечь-жгу*. Реализуются эти чередования с несколькими определенными приставками: *в-/во-, над-/надо-, об-/обо-, от-/ото-, под-/подо-, пред-/предо-, с-/со-,* а также *вз-(вс-)/взо-, воз-(вос-)/возо-, из-(ис-)/изо-, низ-(нис-)/низо-, раз-(рас-)/разо-*. Подробнее (например, правила чередования, примеры и пр.) см. стр. 79-80 в словаре Зализняка.

Помета *ě* или *o* (в столбце *ě,o*)

У глаголов с такой пометой в основе происходит чередование *ě/e* или *o/e*. Подробнее см. стр. 83 в словаре Зализняка.

Цифровые пометы (в столбце РОМЕТУ): (1) → "1" и т.д.

Цифровые пометы показывают, что у глагола представлено некоторое часто встречающееся отклонение от стандартного спряжения. Эти пометы иногда стоят в квадратных скобках, и это означает, что для данного глагола они факультативны. Подробнее см. стр. 81-83 в словаре Зализняка.

1 – касаются ударения. В личных формах Прош. вр. (кроме Ж.р.) и в страдательных причастиях. Прош. вр. (кроме краткой формы Ж.р.) оно падает на один слог левее, чем по общему правилу (практически это означает, что оно падает на приставку). Действительные причастия Прош. вр. и деепричастия Прош. вр. сохраняют ударение инфинитива.

2, 3 – относятся к образованию императива. **2** обозначает, что императив вопреки ожиданиям оканчивается на *-ь(ся), -ьте(сь)* (после гласных на *-й(ся), -йте(сь)*), а не на *-*

и(сь), -ите(сь). **3** обозначает, что повелительное наклонение заканчивается на *-и(сь) // -ь(ся), -ьте(сь)*.

4 – касаются ударения. В действительных причастиях Наст. вр. оно падает на один слог левее, чем ожидалось бы по правилу.

5, 6 – относятся к глаголам класса (НУ). Помета **5** означает, что в форме Ед. ч. М. р. Прош. вр. сохраняется суффикс *-ну-*. Помета **6** означает, что суффикс *-ну-* сохраняется в причастиях и деепричастиях Прош. вр. В сущности, эти пометы помогают нам разделить класс (НУ)^a и (НУ)^b.

7, 8 – касаются ударения. **7** означает, что в страдательных причастиях Прош. вр. оно падает на гласную перед *-нный* (получается *-ённый* и *-áнный*). **8** означает, что в страдательных причастиях Прош. вр. ударение падает на третий слог от конца. Иначе говоря, слоги *-енный* или *-анный* безударные. Отклонение, обратное отклонению **7**.

9 – относится к образованию деепричастий. В роли деепричастия Прош. вр. у глаголов СВ с такой пометой выступает форма на *-а(-я)*, образованная от основы Буд. вр. по правилам, по которым обычно образуются деепричастия Наст. вр. Регулярное образование деепричастий при этом тоже существует, но эти формы устарели и в современном языке неупотребительны.

Пометы и → и (в столбце ,)

Эти пометы обозначают, что образование страд. прич. прош. вр. от данного глагола затруднено (,) или вообще невозможно (,)

Акцентные парадигмы.

Образцы основных схем ударения настоящего времени:

Схема ударения	<i>a/</i>	<i>b/</i>	<i>c/</i>
----------------	-----------	-----------	-----------

Наст./буд. вр., изъяв. накл.	1 л. Ед. ч.		делаю		беру́	ending	пишу́
	2 л. Ед. ч.		делаешь	e	берёшь	stem	пи́шешь
	3 л. Ед. ч.	s	делает	n	берёт	stem	пи́шет
	1 л. Мн. ч.	t	делаем	d	берём	stem	пи́шем
	2 л. Мн. ч.	e	делаете	i	берёте	stem	пи́шете
	3 л. Мн. ч.	m	делают	n	беру́т	stem	пи́шут
Повел. накл.	2 л. Ед. ч.		делай	g	беру́	ending	пишу́
	2 л. Мн. ч.		делайте		беру́те	ending	пи́шите

Образцы основных схем ударения прошедшего времени:

Схема ударения		/a		/b		/c	
Прош. вр., изъяв. накл.	Ед. ч. М. р.		делал	e	бере́г	stem	сорва́л
	Ед. ч. Ж. р.	s	делала	n	берегла́	ending	сорвала́
	Ед. ч. Ср. р.	t	делало	d	берегло́	stem	сорвало́
	Мн. ч.	e	делали	i	берегли́	stem	сорвали́
		m		n			
				g			

Схема ударения		/c''	
Прош. вр., изъяв. накл.	Ед. ч. М. р.	stem / REFL	сорва́лся уст. сорвался
	Ед. ч. Ж. р.	ending	сорвала́сь
	Ед. ч. Ср. р.	ending / stem	сорвало́сь и сорва́лось
	Мн. ч.	ending / stem	сорвали́сь и сорва́лись

В словаре Зализняка акцентные парадигмы Наст./буд. и Прош. вр. даются через знак "/" сразу после номера класса. Мы же в своей базе данных выделили их в два отдельных столбца (PRAES для акцентной парадигмы Наст./буд. вр. и PAST для акцентной парадигмы Прош. вр.).

Примечание:

1. Если указано, что при спряжении ударение падает **на основу**.

Ударение падает на тот же слог основы, что и в инфинитиве. Если это невозможно, так как основа Наст./буд. или Прош. вр. короче (напр., *рисовать* – *рисую* или *тереть* – *тёр*), то на последний слог основы.

2. Если указано, что при спряжении ударение падает **на окончание**.

Ударение падает на первый слог двусложного окончания. Если окончание неслоговое (нуль//ь//й) или если окончание есть нуль//ь//й + *те//ся//тесь*, а ударение по схеме должно падать на него, то ударным оказывается последний слог основы.

Различные особенности:

1. Для глагола *хотеть* существует особая схема ударения Наст. вр., а для глаголов *дать* и *взять* – особая схема ударения Прощ. вр. У Зализняка они обозначаются как *c'* (*хотеть*) и */c'* (*дать* и *взять*). Нам это показалось неудобным: уже есть обозначения *c* и *c''*, так что происходит некоторая путаница. Поэтому мы обозначили эти акцентные парадигмы как *spec.* (в столбцах PRAES и PAST соответственно). Сами схемы устроены следующим образом:

Схема ударения		/c' = spec (PAST)	
Прощ. вр., изъяв. накл.	Ед. ч. М. р.	Stem	да́л
	Ед. ч. Ж. р.	ending	дала́
	Ед. ч. Ср. р.	Stem / ending	да́ло и далó
	Мн. ч.	Stem	да́ли

Схема ударения		c' = spec (PRAES)	
Наст. (буд.) вр., изъяв. накл.	1 л. Ед. ч.	ending	хочу́
	2 л. Ед. ч.	stem	хо́чешь
	3 л. Ед. ч.	stem	хо́чет
	1 л. Мн. ч.	ending	хоту́м
	2 л. Мн. ч.	ending	хоту́те
	3 л. Мн. ч.	ending	хотят
Повел. накл.	2 л. Ед. ч.	ending	хоту́
	2 л. Мн. ч.	ending	хоту́те

2. У различных глаголов есть и другие особенности, связанные с ударением. Если это нерегулярные отклонения, о них сообщается особо. В базе данных вся информация такого рода попала в графу GRAMMAR (там даются целиком формы глагола с ударениями). Если это регулярные отклонения, они помечены в словаре Зализняка цифрами в кружочках. О них см. выше в разделе о значениях помет.

**Приложение 6. Списки бесприставочных и немотивированных
глаголов разных классов.**

Список бесприставочных глаголов непродуктивных классов.

Глаголы	Класс	Кол-во
бить	(И)Й	
вить	(И)Й	
лить	(И)Й	
пить	(И)Й	
шить	(И)Й	5

Глаголы	Класс	Кол-во	Глаголы	Класс	Кол-во
–бегнуть	(НУ)		крепнуть	(НУ)	
блѣкнуть	(НУ)		липнуть	(НУ)	
брюзгнуть	(НУ)		мѣрзнуть	(НУ)	
бухнуть	(НУ)		меркнуть	(НУ)	
–вергнуть	(НУ)		мокнуть	(НУ)	
виснуть	(НУ)		молкнуть	(НУ)	
волгнуть	(НУ)		мякнуть	(НУ)	
воскреснуть	(НУ)		набрякнуть	(НУ)	
–выкнуть	(НУ)		никнуть	(НУ)	
вязнуть	(НУ)		обрыднуть	(НУ)	
вянуть	(НУ)		пахнуть	(НУ)	
гаснуть	(НУ)		промозгнуть	(НУ)	
гибнуть	(НУ)		пухнуть	(НУ)	
глохнуть	(НУ)		разверзнуть	(НУ)	
горкнуть	(НУ)		сипнуть	(НУ)	
грузнуть	(НУ)		склизнуть	(НУ)	
грязнуть	(НУ)		слабнуть	(НУ)	
–двигнуть	(НУ)		слепнуть	(НУ)	
дохнуть	(НУ)		слизнуть	(НУ)	
дрогнуть	(НУ)		сохнуть	(НУ)	
дыхнуть	(НУ)		–стигнуть	(НУ)	
дрябнуть	(НУ)		стынуть	(НУ)	
дряхнуть	(НУ)		сякнуть	(НУ)	
жолкнуть	(НУ)		терпнуть	(НУ)	
жухнуть	(НУ)		тихнуть	(НУ)	
заскорузнуть	(НУ)		–торгнуть	(НУ)	
захряснуть	(НУ)		тускнуть	(НУ)	
зябнуть	(НУ)		тухнуть	(НУ)	
исчезнуть	(НУ)		хрипнуть	(НУ)	
киснуть	(НУ)		чахнуть	(НУ)	60

Глаголы	Класс	Кол-во
(за)чать	/М-/Н	
(на)чать	/М-/Н	
(по)чать	/М-/Н	
(про)пять	/М-/Н	
(рас)пять	/М-/Н	
жать (жму)	/М-/Н	
жать (жну)	/М-/Н	
мять	/М-/Н	8

Глаголы	Класс	Кол-во
мереть	/Р	
переть	/Р	
простереть	/Р	
тереть	/Р	4

Глаголы	Класс	Кол-во
брать	n/s A	
врать	n/s A	
драть	n/s A	
ждать	n/s A	
жрать	n/s A	
звать	n/s A	
лгать	n/s A	
рвать	n/s A	
слать	n/s A	
стлать	n/s A	
ткать	n/s A	11

Глаголы	Класс	Кол-во	Глаголы	Класс	Кол-во
блеять	A		плакать	A	
бормотать	A		плясать	A	
брехать	A		попрать	A	
брызгать	A		прятать	A	
веять	A		пыхать	A	
вопить	A		реготать	A	
воркотать	A		резать	A	
вязать	A		реять	A	
глаголать	A		рокотать	A	
гоготать	A		роптать	A	
грать	A		рукоплескать	A	

грохотать	А		свистать	А	
двигать	А		сеять	А	
деяться	А		сказать	А	
дремать	А		скакать	А	
жаждать	А		скрежетать	А	
заклать	А		–смеять	А	
затеять	А		снискать	А	
искать	А		содеять	А	
казать	А		сосать	А	
каяться	А		стонать	А	
квохтать	А		стрекотать	А	
клеветать	А		сыпать	А	
клекотать	А		таять	А	
клепать	А		тесать	А	
кликать	А		топотать	А	
клокотать	А		топтать	А	
клохтать	А		трепать	А	
колебать	А		трепетать	А	
кудахтать	А		тропотать	А	
лаять	А		хаять	А	
лелеять	А		хвостать	А	
лепетать	А		хлестать	А	
лизать	А		хлобыстать	А	
лопотать	А		хлопотать	А	
мазать	А		хлыстать	А	
маять	А		хохотать	А	
метать	А		цокотать	А	
миропомазать	А		чаять	А	
надеяться	А		чесать	А	
низать	А		чуть	А	
обязать	А		шептать	А	
орать (кричать)	А		щебетать	А	
пахать	А		щекотать	А	
писать	А		щипать	А	91

Глаголы	Класс	Кол-во
орать (пахать)	А // n/s А	1

Глаголы	Класс	Кол-во
алкать	А // АЙ	
глодать	А // АЙ	

колыхать	А // АЙ	
курлыкать	А // АЙ	
махать	А // АЙ	
мурлыкать	А // АЙ	
плескать	А // АЙ	
полоскать	А // АЙ	
рыскать	А // АЙ	
турлыкать	А // АЙ	
тыкать	А // АЙ	
хныкать	А // АЙ	
щепать	А // АЙ	
щипать	А // АЙ	14

Глаголы	Класс	Кол-во
давать	АВАЙ	
–знавать	АВАЙ	
создавать	АВАЙ	
–ставать	АВАЙ	4

Глаголы	Класс	Кол-во
грести	Б	
погрести	Б	
скрести	Б	3

Глаголы	Класс	Кол-во
жить	В	
плыть	В	
слыть	В	3

Глаголы	Класс	Кол-во
беречь	Г-К	
влечь	Г-К	
волочь	Г-К	
жечь	Г-К	
мочь	Г-К	
небредить	Г-К	
облечь	Г-К	
печь	Г-К	
–прядь	Г-К	
–речь	Г-К	
сечь	Г-К	
стеречь	Г-К	

стричь	Г-К	
течь	Г-К	14

Глаголы	Класс	Кол-во
блюсти	Д-Т	
брести	Д-Т	
вести	Д-Т	
гнести	Д-Т	
грясти	Д-Т	
густи	Д-Т	
класть	Д-Т	
красть	Д-Т	
красться	Д-Т	
мести	Д-Т	
мястись	Д-Т	
обрести	Д-Т	
пасть	Д-Т	
плести	Д-Т	
прясть	Д-Т	
рассвести	Д-Т	
цвести	Д-Т	
честь	Д-Т	18

Глаголы	Класс	Кол-во	Глаголы	Класс	Кол-во
бдеть	Е		обидеть	Е	
блестеть	Е		пестреть	Е	
болеть	Е		пыхтеть	Е	
велеть	Е		свербеть	Е	
вертеть	Е		свиристеть	Е	
видеть	Е		свистеть	Е	
висеть	Е		сидеть	Е	
галдеть	Е		сипеть	Е	
глядеть	Е		скорбеть	Е	
гореть	Е		скрипеть	Е	
греметь	Е		смердеть	Е	
гудеть	Е		смотреть	Е	
дудеть	Е		сопеть	Е	
звенеть	Е		тарахтеть	Е	
зреть	Е		терпеть	Е	
зудеть	Е		трухтеть	Е	
кипеть	Е		храпеть	Е	
кишеть	Е		хрипеть	Е	
коптеть	Е		хрустеть	Е	

корпеть	Е		шелестеть	Е	
кряхтеть	Е		шипеть	Е	
лететь	Е		шуметь	Е	45
лицезреть	Е				

Глаголы	Класс	Кол-во	Глаголы	Класс	Кол-во
бояться	ЖА		мурчать	ЖА	
бренчать	ЖА		мчать	ЖА	
брюзжать	ЖА		мычать	ЖА	
бурчать	ЖА		пищать	ЖА	
верезжать	ЖА		рычать	ЖА	
верещать	ЖА		сверчать	ЖА	
визжать	ЖА		скворчать	ЖА	
ворчать	ЖА		слышать	ЖА	
держать	ЖА		стоять	ЖА	
дребезжать	ЖА		стучать	ЖА	
дрожать	ЖА		торчать	ЖА	
дышать	ЖА		трещать	ЖА	
жужжать	ЖА		урчать	ЖА	
журчать	ЖА		фурчать	ЖА	
звучать	ЖА		фырчать	ЖА	
кричать	ЖА		шуршать	ЖА	
лежать	ЖА		ячать	ЖА	35
молчать	ЖА				

Глаголы	Класс	Кол-во
везти	З-С	
грызть	З-С	
лезть	З-С	
нести	З-С	
отверзть	З-С	
пасти	З-С	
ползти	З-С	
трясти	З-С	8

Глаголы	Класс	Кол-во
гнить	Й	
дуть	Й	
обуть	Й	
почить	Й	
разуть	Й	5

Глаголы	Класс	Кол-во
деть	Н	
стать	Н	
–стрячь	Н	
стыть	Н	4

Глаголы	Класс	Кол-во
–бороть	О	
колоть	О	
полоть	О	
пороть	О	4

Глаголы	Класс	Кол-во
выть	ОЙ	
крыть	ОЙ	
мыть	ОЙ	
ныть	ОЙ	
рыть	ОЙ	5

Список неправильных бесприставочных глаголов.

Глаголы	Класс	Глаголы	Класс
дать	ANOMAL	сесть	IRREG (Д-Т)
есть	ANOMAL	реветь	IRREG (Е)
надоесть	ANOMAL	хотеть	IRREG (Е)
создать	ANOMAL	бежать	IRREG (ЖА)
гнать	IRREG (n/s A)	родить	IRREG (И)
спать	IRREG (n/s A)	читать	IRREG (И)
–шибить	IRREG (Б)	брить	IRREG (Й)
лечь	IRREG (Г-К)	петь	IRREG (Й)
толочь	IRREG (Г-К)	взять	IRREG (Й/М-НИМ)
быть	IRREG (Д-Т)	принять	IRREG (Й/М-НИМ)
ехать	IRREG (Д-Т)	клясть	IRREG (Н)
идти (–йти)	IRREG (Д-Т)	молоть	IRREG (О)
расти	IRREG (Д-Т)		

Список бесприставочных глаголов класса ЕЙ.

Глаголы	Класс	Кол-во	Глаголы	Класс	Кол-во
алеть	ЕЙ		одряблеть	ЕЙ	
багроветь	ЕЙ		окосеть	ЕЙ	
багрянеть	ЕЙ		олонеть	ЕЙ	
беднеть	ЕЙ		опаскудеть	ЕЙ	

белеть	ЕЙ		ополоуметь	ЕЙ	
беременеть	ЕЙ		опротиветь	ЕЙ	
благоговеть	ЕЙ		осопливеть	ЕЙ	
бледнеть	ЕЙ		остеклеть	ЕЙ	
богатеть	ЕЙ		осточертеть	ЕЙ	
болеть	ЕЙ		оторопеть	ЕЙ	
брюхатеть	ЕЙ		очертенеть	ЕЙ	
буреть	ЕЙ		очертеть	ЕЙ	
вдоветь	ЕЙ		ошепеляветь	ЕЙ	
ведренеть	ЕЙ		паршиветь	ЕЙ	
веселеть	ЕЙ		пестреть	ЕЙ	
ветренеть	ЕЙ		–печатлеть	ЕЙ	
ветреть	ЕЙ		пламенеть	ЕЙ	
вечереть	ЕЙ		плесневеть	ЕЙ	
взрослеть	ЕЙ		плешиветь	ЕЙ	
–виднеть	ЕЙ		плотнеть	ЕЙ	
владеть	ЕЙ		подлеть	ЕЙ	
влажнеть	ЕЙ		пожидеть	ЕЙ	
воденеть	ЕЙ		полнеть	ЕЙ	
вождедель	ЕЙ		посолиднеть	ЕЙ	
волосатеть	ЕЙ		постылеть	ЕЙ	
вшиветь	ЕЙ		потеть	ЕЙ	
глазеть	ЕЙ		пошлеть	ЕЙ	
глупеть	ЕЙ		праветь	ЕЙ	
говеть	ЕЙ		преть	ЕЙ	
голубеть	ЕЙ		присмиреть	ЕЙ	
горбатеть	ЕЙ		прозреть	ЕЙ	
греть	ЕЙ		простеть	ЕЙ	
грубеть	ЕЙ		прочнеть	ЕЙ	
грузнеть	ЕЙ		прыщаветь	ЕЙ	
грустнеть	ЕЙ		псоветь	ЕЙ	
грязнеть	ЕЙ		пустеть	ЕЙ	
густеть	ЕЙ		пьянеть	ЕЙ	
деревенеть	ЕЙ		радеть	ЕЙ	
дернеть	ЕЙ		разуметь	ЕЙ	
дешеветь	ЕЙ		рдеть	ЕЙ	
добреть	ЕЙ		редеть	ЕЙ	
довлеть	ЕЙ		ржаветь	ЕЙ	
дряхлеть	ЕЙ		робеть	ЕЙ	
дуреть	ЕЙ		роговеть	ЕЙ	
дурнеть	ЕЙ		розоветь	ЕЙ	
дюжеть	ЕЙ		румянеть	ЕЙ	
жаднеть	ЕЙ		русеть	ЕЙ	
жалеть	ЕЙ		ручнеть	ЕЙ	
желтеть	ЕЙ		рыжеть	ЕЙ	

жиреть	ЕЙ		рыхлеть	ЕЙ	
жирнеть	ЕЙ		рябеть	ЕЙ	
заглянцеветь	ЕЙ		саднеть	ЕЙ	
задубенеть	ЕЙ		сатанеть	ЕЙ	
задубеть	ЕЙ		свежеть	ЕЙ	
закоренеть	ЕЙ		светлеть	ЕЙ	
заштилеть	ЕЙ		свирепеть	ЕЙ	
звереть	ЕЙ		седеть	ЕЙ	
здороветь	ЕЙ		сереть	ЕЙ	
зеленеть	ЕЙ		сиветь	ЕЙ	
злеть	ЕЙ		сизеть	ЕЙ	
золотеть	ЕЙ		синеть	ЕЙ	
зреть	ЕЙ		сиротеть	ЕЙ	
иметь	ЕЙ		сквернеть	ЕЙ	
индеветь	ЕЙ		скудеть	ЕЙ	
каменеть	ЕЙ		скучнеть	ЕЙ	
–колеть	ЕЙ		слабеть	ЕЙ	
косматеть	ЕЙ		смелеть	ЕЙ	
коснеть	ЕЙ		сметь	ЕЙ	
костенеть	ЕЙ		смирнеть	ЕЙ	
коченеть	ЕЙ		смуглеть	ЕЙ	
краснеть	ЕЙ		советь	ЕЙ	
криветь	ЕЙ		соловеть	ЕЙ	
кروаветь	ЕЙ		солодеть	ЕЙ	
кровенеть	ЕЙ		спесиветь	ЕЙ	
круглеть	ЕЙ		спеть	ЕЙ	
крупнеть	ЕЙ		стареть	ЕЙ	
куржаветь	ЕЙ		стекленеть	ЕЙ	
курчаветь	ЕЙ		стервенеть	ЕЙ	
леветь	ЕЙ		столбенеть	ЕЙ	
леденеть	ЕЙ		строптиветь	ЕЙ	
лиловеть	ЕЙ		студенеть	ЕЙ	
лиственеть	ЕЙ		суроветь	ЕЙ	
лубенеть	ЕЙ		сыреть	ЕЙ	
лысеть	ЕЙ		сытеть	ЕЙ	
люднеть	ЕЙ		твердеть	ЕЙ	
лютеть	ЕЙ		темнеть	ЕЙ	
матереть	ЕЙ		теплеть	ЕЙ	
махроветь	ЕЙ		тлеть	ЕЙ	
мелеть	ЕЙ		толстеть	ЕЙ	
мерзеть	ЕЙ		травенеть	ЕЙ	
мертветь	ЕЙ		трезветь	ЕЙ	
милеть	ЕЙ		трухлеть	ЕЙ	
млеть	ЕЙ		трухляветь	ЕЙ	
молодеть	ЕЙ		тупеть	ЕЙ	

мохнатеть	ЕЙ		тускнеть	ЕЙ	
мрачнеть	ЕЙ		тучнеть	ЕЙ	
мутнеть	ЕЙ		тяготеть	ЕЙ	
–мшеть	ЕЙ		тяжелеть	ЕЙ	
наглеть	ЕЙ		угрюметь	ЕЙ	
натореть	ЕЙ		уметь	ЕЙ	
неметь	ЕЙ		умнеть	ЕЙ	
новеть	ЕЙ		успеть	ЕЙ	
обалдеть	ЕЙ		уцелеть	ЕЙ	
обезводеть	ЕЙ		хаметь	ЕЙ	
обезволеть	ЕЙ		хилеть	ЕЙ	
обезголосеть	ЕЙ		хиреть	ЕЙ	
обезденежить	ЕЙ		хладеть	ЕЙ	
обезземелеть	ЕЙ		хмелеть	ЕЙ	
обеззубеть	ЕЙ		холодеть	ЕЙ	
обезлесеть	ЕЙ		холоднеть	ЕЙ	
обезлошадеть	ЕЙ		хорошеть	ЕЙ	
обезлюдеть	ЕЙ		храбреть	ЕЙ	
обезматочеть	ЕЙ		хрометь	ЕЙ	
обезножить	ЕЙ		худеть	ЕЙ	
обезрыбеть	ЕЙ		цепенеть	ЕЙ	
обезуметь	ЕЙ		червиветь	ЕЙ	
обескроветь	ЕЙ		чернеть	ЕЙ	
обеспамятеть	ЕЙ		черствовать	ЕЙ	
обесплодеть	ЕЙ		чуметь	ЕЙ	
обессилеть	ЕЙ		шалеть	ЕЙ	
оболванеть	ЕЙ		шелудиветь	ЕЙ	
обуглеть	ЕЙ		шершаветь	ЕЙ	
обутреть	ЕЙ		юнеть	ЕЙ	
оголеть	ЕЙ		ядренеть	ЕЙ	
одолеть	ЕЙ		яловеть	ЕЙ	
одревеснеть	ЕЙ		яснеть	ЕЙ	252

Список немотивированных глаголов некоторых продуктивных классов.

(в основном взят из Академической грамматики)

АЙ-класс

В АЙ-класс входят глаголы на *-ать (-ять), -ивать (-ывать), -вать, -ничать, -ичать*, мотивированные именами, междометиями, глаголами, а также немотивированные глаголы:

блуждать, бодать, брезгать, валять, ведать, верстать, вещать, вихлять, влиять, ворочать, глотать, гулять, делать, дёргать, дрыгать, дрызгать, ерзать, зевать, зиять, знать, играть, издеваться, икать, калякать, карать, качать, кивать, кидать, ковылять, ковырять, корнать, кромсать, кумекать, купать, кусать, кутать, кушать, лакать, ласкать, латать, линять, лобзать, ломать, лопать, лягать, макать, марать, мелькать, мешать, мигать, мотать, намереваться, недоумевать, нырять, обонять, обуревать, обувать, падать, пачкать, паять, пенять, питать, пихать, пичкать, плутать, подозревать, поймать, почивать, притязать, прыскать, прядать, пылать, ронять, ругать, ручаться, рыкать, сверкать, сиять, скучать, созерцать, сомневаться, спрягать, стенать, стрелять, строговать, стряпать, стучать, терзать, терять, тика́ть, тиска́ть, тискать, толкать, тягать, тягаться, увещевать, уповать, хвастать, хватать, хворать, хлебать, читать, чихать, шатать, швырять, шнырять, щупать

Всего их 107. Интересно, что *знать* – единственный в этом классе глагол с односложной основой.

ЕЙ-класс

В ЕЙ-класс входят глаголы, мотивированные существительными и прилагательными, а также группа немотивированных глаголов:

греть, зреть (спеть), млеть, преть, рдеть, сметь, спеть, тлеть, болеть, владеть, возделеть, говеть, доветь, жалеть, иметь, -колеть (о-, пере-), коченеть, натореть, одолеть, прозреть, саднеть, уметь, успеть, хиреть, цепенеть

Всего их 25.

ОВА-класс

К ОВА-классу относятся мотивированные суффиксальные глаголы с морфами *-ова(ть), -ирова(ть), -изирова(ть), -изова(ть), -ствова(ть), -ествова(ть)*, сложные глаголы со связным опорным компонентом *-фицирова(ть)* и немотивированные глаголы:

1. с многосложной основой наст. вр. (напр., *аккредитировать, ассигновать, аттестовать, баловать* и т. д.)
2. с односложной основой наст. вр.: *блевать, жевать, клевать, ковать, плевать, сновать, совать* (эти 7 глаголов выделяются в отдельную группу и имеют особую

акцентную парадигму Наст. вр.: в настоящем времени ударение у них падает на окончание, в то время как у всех прочих глаголов ОВА-класса – на основу).

Количество немотивированных глаголов ОВА-класса с многосложной основой Наст. вр. пока не посчитано.

НУ-класс

В НУ-класс входят мотивированные глаголы с суффиксами *-ну(ть)* и *-ану(ть)*, имеющие значение одноактности действия (напр., *двинуть, кольнуть, пихнуть, прыгнуть, резануть*), а также ряд немотивированных глаголов:

1. с односложной основой: *гнуть, льнуть*
2. с неодносложной основой: *вернуть, -вихнуть (с-, вы-), вспыхнуть, всхлипнуть, вынуть, -гинуть (с-), грянуть, запнуться, кануть, минуть, обернуть, обмануть, очнуться, -пахнуть (рас-, за-), поперхнуться, прикорнуть, прынуть, -прынуть (вос-, от-), рехнуться, -ринуть (низ-, от-), ринуться, рухнуть, -стегнуть (за-, от-, при-, рас-), тонуть, улепетнуть, хлынуть, шелохнуться*

Всего 29 (если, как мы делали это в списках бесприставочных глаголов, считать те глагольные основы, которые сейчас представлены только в сочетании с префиксами, за один глагол). Большая часть немотивированных глаголов не обладает значением одноактности.

И-класс

К И-классу относятся глаголы, мотивированные именами, и большая группа немотивированных глаголов.