ЧАСТОТНОСТИ РАЗЛИЧНЫХ ГРАММАТИЧЕСКИХ ХАРАКТЕРИСТИК И ОКОНЧАНИЙ У СУЩЕСТВИТЕЛЬНЫХ РУССКОГО ЯЗЫКА

Слюсарь Н. А. (slioussar@gmail.com)[1,2],
Самойлова М. В. (ajmi@yandex.ru)[2]

[1]НИУ ВШЭ, Москва, [2]СПбГУ, Санкт-Петербург, Россия

Ключевые слова: русский язык, род, число, падеж, одушевленность, парадигмы существительных, частотность

FREQUENCIES OF DIFFERENT GRAMMATICAL FEATURES AND INFLECTIONAL AFFIXES IN RUSSIAN NOUNS

Slioussar N. A. (slioussar@gmail.com)[1, 2],
Samoilova M. V. (ajmi@yandex.ru)[2]

[1]HSE, Moscow, [2]St. Petersburg State University, St. Petersburg, Russia

Most researchers working with linguistic data sooner or later have to appeal to frequency. In many cases, an *ad hoc* comparison can be done: e. g. we can find out whether Dative is less frequent than Prepositional using one of the existing Russian corpora. However, creating a systematic database where frequencies of different grammatical characteristics are estimated on one corpus sample is a better solution. This way, we can see from the very start how the contrast we are interested in depends on other grammatical properties.

We present a database that contains information about frequencies of different grammatical features and inflectional affixes in Russian nouns (http://www.slioussar.ru/freqdatabase.html). It was created on the basis of the grammatically disambiguated subcorpus of the Russian National Corpus (http://www.ruscorpora.ru). We analyzed the following grammatical categories: gender, number, case, animacy (by themselves and in various combinations), looked at these categories in different declensions, and also at the distribution of different inflectional affixes. Such data are crucial for many theoretical and experimental approaches, especially for usage-based ones. They may be useful not only per se, but also in solving auxiliary problems: for linguists, psychologists and other cognitive scientists choosing linguistic stimuli for their experiments.

Keywords: Russian, number, gender, case, animacy, nominal inflectional paradigms, frequency

1. Введение

В данной работе была поставлена цель получить сведения о частотности различных грамматических характеристик существительных русского языка, опираясь на подкорпус Национального корпуса русского языка (НКРЯ) со снятой неоднозначностью (http://www.ruscorpora.ru). Одной из задач было определить, насколько частотны формы существительных разного рода, в разных числах и падежах, одушевленных и неодушевленных, как эти характеристики зависят от словоизменительных парадигм (склонения и типа основы) и как они коррелируют друг с другом. Вторая задача заключалась в том, чтобы определить частотность форм с различными окончаниями (в зависимости от падежа, числа, рода и склонения и вне зависимости от них). Собранные сведения о частотности были объединены в небольшую базу данных, предварительная версия которой доступна в интернете (http://www.slioussar.ru/freqdatabase.html).

Очевидно, что, если нужно сравнить, скажем, частотность двух падежей, несложно сделать запрос в НКРЯ, не пользуясь никакой базой. База нужна для того, чтобы получить общую картину (например, частотности интересующих падежей на фоне всех падежей), а также иметь возможность впоследствии включить в сравнение новые факторы, скажем, число или одушевленность. Ведь, хотя различные системы автоматического анализа, разработанные для русского языка, основаны на такого рода статистике, соответствующая информация пока не представлена в открытом доступе в обобщенном виде. Также важно отметить, что существует несколько проектов, посвященных исследованию частотности падежных и других форм в русском языке (например, Копотев 2008; Lyashevskaya 2013). Однако они направлены прежде всего на описание особенностей парадигм отдельных слов.

Сведения о частотности грамматических форм с учетом разных словоизменительных классов лексем и о частотности окончаний необходимы для целого ряда теоретических и экспериментальных лингвистических исследований, в особенности для всего спектра подходов, ориентированных на употребление, т. н. *usage-based* (Baayen 2003; Bybee 2006; Dressler 1985; Milin et al. 2009; Moscoso del Prado Martín et al. 2004 и мн. др.), а также для любых моделей, описывающих ментальный лексикон носителя: к какому бы направлению они ни относились, частотность всегда играет в них ту или иную важную роль. Причем сведения подобного рода могут быть востребованы как сами по себе (например, исследуя, как представлены в ментальном лексиконе грамматические категории рода, числа, падежа, важно знать частотность различных граммем), так и для решения вспомогательных задач, скажем, при подборе стимулов для психолингвистических экспериментов.

Проект осуществляется при поддержке гранта РГНФ №14-04-12034.

2. Несколько слов о роли частотности в различных моделях ментального лексикона

Сколько-нибудь полный обзор того, какую роль играет частотность в различных моделях ментального лексикона, потребовал бы отдельной большой статьи, поэтому здесь мы лишь вкратце коснемся этой темы. В исследованиях ментального лексикона значительное место занимает дискуссия между сторонниками односистемного и двусистемного подходов к морфологически сложным словам. Сторонники двусистемного подхода (например, Clahsen 1999; Marslen-Wilson, Tyler 1997; Pinker 1991, 1999; Pinker, Prince 1988; Ullman 2004) делят формы на правильные и неправильные. Первые образуются по правилу при порождении и подвергаются морфологическому анализу при восприятии, то есть расчленяются на отдельные морфемы. Вторые же хранятся в ассоциативной памяти целиком.

Согласно односистемному подходу (например, MacWhinney, Leinbach 1991; McClelland, Patterson 2002; Plunkett, Marchman 1993; Rummelhart, McClelland 1986), все формы обрабатываются единым механизмом. Они порождаются и воспринимаются за счет аналогии с другими формами. Этот подход отрицает психолингвистическую реальность символических правил и эксплицитного морфологического анализа. Таким образом, изначально представители односистемного подхода полагали, что частотность играет решающую роль для обработки любых форм, а представители двусистемного подхода — что только для обработки неправильных. Тем не менее, в результате ряда экспериментов (например, Alegre, Gordon, 1999), в основе которых лежали как раз манипуляции с частотностью словоформ и лемм, даже сторонники двусистемного подхода согласились с тем, что наиболее частотные регулярные формы представлены в памяти не только в виде отдельных морфем, но и целиком.

Х. Баайен и его последователи, которых можно отнести к односистемному подходу, продемонстрировали, что для восприятия словоформ имеет принципиальное значение не только их собственная частотность и частотность леммы, но и количество форм в парадигме, к которой они относится, частотность этих форм, а также частотность этой парадигмы по сравнению с другими парадигмами (например, Baayen et al. 1997, 2003; Kostič 1991, 1995; Milin et al. 2009). Для описания этих закономерностей были использованы методы теории информации. Таким образом, для проведения дальнейших исследований на материале русского языка представляется необходимым собрать соответствующую информацию о русских существительных.

3. Описание базы данных

База данных создавалась следующим образом. В подкорпусе НКРЯ со снятой неоднозначностью были собраны сведения о частотности форм существительных с различными окончаниями. При этом учитывались такие параметры, как род, число, падеж, одушевленность и тип склонения, а также тип основы (подробнее об этом рассказано в разделе 3.2). Вся собранная информация

и сделанные затем на ее основании расчеты помещались в таблицы в файл «freqdatabase.xlsx». Его текущая версия доступна по адресу http://www.sliousar.ru/freqdatabase.html. В базу данных не вошли аббревиатуры, неизменяемые существительные и существительные адъективного склонения (их окончания совпадают с окончаниями прилагательных, поэтому становится сложно говорить о частотности окончаний). Некоторые другие исключения (всегда крайне немногочисленные) оговорены в разделе 3.3 и, если речь о совсем частных случаях, на листах с исходными данными.

3.1. Несколько слов о парадигмах русских существительных

А. А. Зализняк (1977) в «Грамматическом словаре русского языка» описывает склонение существительных так. Для каждого рода есть набор основных окончаний. По сути, для м. и ср. р. это парадигма 1 скл., для женского — парадигма 2 скл.[1] Этот набор существует в двух вариантах: первые обычно используются для основ на твердые согласные, вторые — для основ на мягкие (см. Таблицы 1–2). У слов с основами на *-г/к/х и -ш/ж/ч/щ/ц* (заднеязычные, шипящие и *ц*) окончания «твердого» и «мягкого» варианта смешаны, причем основы на *-ц* образуют особую группу, так как у них почти все окончания из «твердого» набора. Имеют свои особенности и основы на *-й*, но окончания у них из того же набора, что используется для других мягких согласных. Кроме того, есть слова м. р., относящиеся ко 2 скл.

Особняком стоят слова ж. р., оканчивающиеся на *-ь* (3 скл.), у которых свой набор окончаний (см. Таблицу 3), а также т. н. разносклоняемые: слова ср. р., оканчивающиеся на *-мя*, и слово *путь*. В файле «freqdatabase.xlsx» они обозначены как *irreg.* Наконец, есть слова адъективного склонения, которые не учтены в базе.

Таблица 1. Основной набор окончаний для слов м., ж. и ср. р. в ед. ч.

		М. р.		**Ср. р.**		**Ж. р.**		**Особенности**
		«тв.» вар.	**«мягк.» вар.**	**«тв.» вар.**	**«мягк.» вар.**	**«тв.» вар.**	**«мягк.» вар.**	
Им.		*0*	*0 (-ь/-й)*	*о*	*е*	*а*	*я*	
Род.		*а*	*я*	*а*	*я*	*ы*	*и*	
Дат.		*у*	*ю*	*у*	*ю*	*е*	*е*	
Вин.	одуш.	*а*	*я*	*о*	*е*	*у*	*ю*	
	неодуш.	*0*	*0 (-ь/-й)*					
Тв.		*ом*	*ем*	*ом*	*ем*	*ой*	*ей*	В ж. р. архаичные варианты *-ою/ею*.
Предл.		*е*	*е*	*е*	*е*	*е*	*е*	

[1] Нумерация склонений дается по «Русской грамматике» (1980), т. е. слово 1 скл. — это, например, *стол*.

Таблица 2. Основной набор окончаний для слов м., ж. и ср. р. в мн. ч.

		М. р.		**Ср. р.**		**Ж. р.**		**Особенности**
		«тв.» вар.	**«мягк.» вар.**	**«тв.» вар.**	**«мягк.» вар.**	**«тв.» вар.**	**«мягк.» вар.**	
Им.		*ы*	*и*	*а*	*я*	*ы*	*и*	В м. р. *-а/я*, *-е*[2], в ср. р. *-и*.
Род.		*ов*	*ей*	*0*	*0 (-ь/-й) //ей*	*0*	*0 (-ь/-й) //ей*	Много вариации, но доп. окончание одно: *-ев*.
Дат.		*ам*	*ям*	*ам*	*ям*	*ам*	*ям*	
Вин.	одуш.	*ов*	*ей*	*0*	*0 (-ь/-й) //ей*	*0*	*0 (-ь/-й) //ей*	
	неодуш.	*ы*	*и*	*а*	*я*	*ы*	*и*	В м. р. *-а/я*, в ср. р. *-и*.
Тв.		*ами*	*ями*	*ами*	*ями*	*ами*	*ями*	
Предл.		*ах*	*ях*	*ах*	*ях*	*ах*	*ях*	

Таблица 3. Окончания 3 скл.

	Ед. ч.	**Мн. ч.**
Им.	*0 (-ь)*	*и*
Род.	*и*	*ей*
Дат.	*и*	*ям/ам*
Вин.	*0 (-ь)*	*и*
Тв.	*-ью*	*ями/ами*
Предл.	*и*	*ях/ах*

Подавляющее большинство остальных особенностей склонения касается изменения основы и иногда нестандартного использования окончаний из основного набора, но не использования каких-то новых окончаний. Самые частотные исключения обозначены ниже (есть еще единичные слова-исключения типа *дитя* и пр.). Огромный пласт очень важной информации, которой мы никак не касаемся, — акцентные парадигмы.

3.2. Как были собраны исходные данные

В идеале нас интересует частотность тех или иных грамматических характеристик и окончаний в рамках определенной парадигмы. Однако у русских существительных очень много различных особенностей в склонении, и учесть их все, т. е. посчитать всё по отдельности для всех возможных типов, представляется слишком сложным. Мы решили остановиться на полпути, сделав расчеты для основ на твердые, мягкие и шипящие и заднеязычные согласные.

---

[2] Только у одушевленных существительных типа *крестьянин — крестьяне*.

Мы обратились к подкорпусу НКРЯ со снятой грамматической омонимией. Первичный сбор информации осуществлялся в октябре — декабре 2013 г. В выбранный нами подкорпус входят почти 6 миллионов слов из примерно 230 миллионов, т. е. достаточно много, хотя это и небольшой процент от всего корпуса.

Для сбора данных использовался лексико-грамматический поиск. В графе «грамматические признаки» мы выбирали, например: сущ., м. р., им. п., мн. ч., одуш. (последнее позволяет сравнить падежи у одушевленных и неодушевленных существительных, не говоря уже о морфологической роли этой категории), затем исключали неизменяемые слова и сокращения. Запрос в результате выглядел так: *S,nom,pl,m,anim -abbr -0*. В графе «слово» ставили, например: *("*ки "|"*ги"|"*хи"|"*ши"|"*жи"|"*чи"|"*щи") -*а,* т. е. формы на *-ки, -ги, -хи, -ши,- -жи, -чи, -щи*, но не от слов на *-а*. Последнее позволяет исключить слова м. р. 2 скл., например, *юноши* или *скряги*. В тех падежах, где у одних и тех же основ есть два варианта окончаний (скажем, *-ой/-ей* у основ на *-ц* в зависимости от ударения), оба варианта были обсчитаны отдельно, чтобы можно было посмотреть, какой частотней. Следует признать, что процедуру сбора данных можно бы было сделать более технологичной, однако на итоговый результат это не влияет.

На основании собранных таким образом данных были произведены все дальнейшие расчеты. Однако все исходные данные, а также информация о запросах включены в файл «freqdatabase.xlsx», чтобы в случае необходимости было проще произвести какие-то дополнительные вычисления или как-то иначе сгруппировать исходные данные (например, разделить основы на шипящие и заднеязычные). Заметим также, что некоторые запросы, в частности, приведенный выше, оказались для системы поиска в НКРЯ слишком длинными. В таких случаях система выдавала ответ «Сервис временно недоступен». Мы разбивали такие запросы надвое или натрое, а потом складывали полученные числа, но в файле «freqdatabase.xlsx» этого не указывали.

3.3. Некоторые особенности НКРЯ и того, как мы использовали представленную там информацию

Рассматривая собранные нами данные, надо иметь в виду следующие особенности НКРЯ. Во-первых, поиск в НКРЯ не учитывает букву *ё*. Во-вторых, в НКРЯ кроме мужского, женского и среднего выделяется общий род (около 5 тысяч примеров). К нему отнесены одушевленные существительные 2 скл. типа *убийца* (около 3,5 тысяч примеров) и некоторые другие менее интересные для нас слова вроде несклоняемых фамилий (около 1,5 тысяч). Мы его пока не учитывали. В-третьих, кроме основных шести падежей в НКРЯ выделяются еще несколько, перечисленные в Таблице 4.

Также важно отметить, что в НКРЯ допущено некоторое количество ошибок при разметке материала. Это вносит определенную погрешность в полученные нами результаты. Однако, как мы покажем в разделе 3.4 на одном примере, погрешность эта очень небольшая (кроме того, следует учесть, что наши данные по определению являются приблизительными, так как получены

на материале определенной корпусной выборки). Иногда, впрочем, ошибки начинают играть более существенную роль. Например, изначально мы нашли сколько-то форм, определенных как неодушевленные существительные м. р. на *-а* и *-я*. При ближайшем рассмотрении оказалось, что все эти случаи — результат различных ошибок, в основном формы типа *методами*, которые привязаны к двум леммам: *метод* (сущ. м.р.) и *метода* (сущ. на *-а*). Мы надеемся, что нам удалось избавиться от всех серьезных проблем такого рода, но, конечно, не можем быть в этом уверенными.

Таблица 4. Падежи, выделяемые в НКРЯ в дополнение к основным шести

Звательный	Формы типа *мам, боже.* 659 существительных в ед. ч.: существительные м. и ж. р. 2 скл. (а также шесть имен типа *Паш*, спорно отнесенных к общему роду) и архаичные формы. Пока не учитывали.
Родительный 2	Формы типа *(стакан) чаю.* Сущ. м.р. 1 скл. в ед. ч., а также почему-то форма *пол(у)ночи* (23 шт.).
Винительный 2	Формы типа (*идти в) солдаты*. 565 существительных м., ж. и общ. р. 1, 2, 3 скл. во мн. ч. Пока не учитывали.
Предложный 2	Формы типа *(в) лесу, (в) сети́*. Существительные м. р. 1 скл. и ж. р. 3 скл. в ед. ч., а также почему-то аббревиатура *гг.* (21 шт.).
Счетная форма	Формы типа *(три) стола.* 676 существительных м. р. 1 скл. Пока не учитывали.

3.4. «Контрольный замер»

Мы считаем важным, что все наши расчеты были произведены на одном и том же массиве данных — это позволяет без каких бы то ни было оговорок сравнивать их между собой. Тем не менее, мы решили провести нечто вроде «контрольного замера», посчитав частотность некоторых грамматических характеристик другим способом. В подкорпусе НКРЯ со снятой омонимией мы собрали сведения о количестве форм одушевленных и неодушевленных существительных разного рода и о количестве форм одушевленных и неодушевленных существительных в разных числах и падежах. В расчеты вошли все формы, классифицированные в НКРЯ как существительные, исключая аббревиатуры и неизменяемые, но включая адъективное склонение. Окончания, склонения и типы основ не учитывались.

Сумма всех форм, вошедших в наши основные расчеты, — 1 544 051, а в контрольные — 1 646 295 (данные по роду) и 1 647 107 (данные по числу и падежу). При этом, как можно удостовериться в файле «freqdatabase.xlsx», распределение форм по грамматическим категориям практически совпадает. Различия между выборками связаны с тем, что в основных расчетах мы давали намного более подробную характеристику форм, и в результате было исключено некоторое их количество (как ошибки, так и подходящие формы, у которых

по ошибке не проставлены те или иные характеристики). Кроме того, как было сказано в разделе 2.3, мы исключили из поиска некоторые типы форм. В результате получается, например, что, когда мы делали дополнительные расчеты по числу и падежу, мы не брали звательный падеж, а в расчеты по роду эти формы вошли. Именно поэтому нам кажется важным, что все наши основные расчеты сделаны на одной выборке.

3.5. Как устроен файл с базой данных

Листы с исходными данными в файле «freqdatabase.xlsx» помещены в конец и выделены серым. Листы с расчетами выделены голубым (там, где расчеты сделаны с учетом склонения, использован светло-голубой). Зеленым отмечен лист с контрольными расчетами (см. раздел 3.4). Мы будем стремиться к тому, чтобы вся информация в файле приводилась по-русски и по-английски, окончания и прочее давались кириллицей и в транслитерации. Однако пока английский язык и транслитерацию можно найти только на листах с основными результатами. Там, где это возможно, мы используем общеупотребительные обозначения грамматических категорий латинского происхождения.

4. Несколько примеров использования содержащейся в базе данных информации

Во введении мы упомянули о том, что для простых сравнений — например, чтобы сопоставить частотность двух падежей — нет необходимости пользоваться базой данных, можно просто задать запрос в НКРЯ. Покажем, как база данных может быть полезна для более сложных расчетов, на примере сравнения дательного и предложного падежей. Как видно из Таблиц 5 и 6, в целом предложный падеж в два раза частотней дательного, однако картина усложняется, если учитывать одушевленность и число. В частности, дательный падеж частотней предложного для одушевленных существительных, особенно в единственном числе.

Таблица 5. Частотности различных падежей с учетом и без учета одушевленности

	Одуш.	Неодуш.	Всего	Одуш.	Неодуш.	Всего
Им.	216 713	254 188	470 901	56,5%	21,9%	30,5%
Род.	75 072	317 552	392 624	19,6%	27,4%	25,4%
Дат.	24 326	51 399	75 725	6,3%	4,4%	4,9%
Вин.	35 478	264 395	299 873	9,3%	22,8%	19,4%
Твор.	27 540	124 973	152 513	7,2%	10,8%	9,9%
Предл.	4 377	148 038	152 415	1,1%	12,8%	9,9%
Всего	383 506	1 160 545	1 544 051	100,0%	100,0%	100,0%

Таблица 6. Частотности различных падежей в зависимости от числа

		Одуш.	Неодуш.	Всего	Одуш.	Неодуш.	Всего
Ед.	Им.	184345	201250	385595	48,1%	17,3%	25,0%
Ед.	Род.	50312	227715	278027	13,1%	19,6%	19,0%
Ед.	Дат.	18982	38985	57967	4,9%	3,4%	3,8%
Ед.	Вин.	26702	211476	238178	7,0%	18,2%	15,4%
Ед.	Твор.	20646	95746	116392	5,4%	8,3%	7,5%
Ед.	Предл.	2961	120761	123722	0,8%	10,4%	8,0%
Всего		**303948**	**895933**	**1199881**	**79,3%**	**77,2%**	**77,7%**
Мн.	Им.	32368	52938	85306	8,4%	4,6%	5,5%
Мн.	Род.	24760	89837	114597	6,5%	7,7%	7,4%
Мн.	Дат.	5344	12414	17758	1,4%	1,1%	1,2%
Мн.	Вин.	8776	52919	61695	2,3%	4,6%	4,0%
Мн.	Твор.	6894	29227	36121	1,8%	2,5%	2,3%
Мн.	Предл.	1416	27277	28693	0,4%	2,4%	1,9%
Всего		**79558**	**264612**	**344170**	**20,7%**	**22,8%**	**22,3%**

Теперь предположим, что нас интересует грамматическая омонимия и падежный синкретизм, которые активно исследуются в рамках различных теоретических и экспериментальных подходов к морфологии (например, Baerman et al. 2005). У дательного и предложного падежа во втором склонении одинаковое окончание *-е*. Но можно ли, например, говорить о том, что распределение существительных между дательным и предложным падежом во втором склонении и в целом похожи? Данные в Таблице 7 (вырезанные из более большой таблицы в файле, охватывающей все падежи), которые можно сравнить с Таблицей 6, позволяют ответить на этот вопрос положительно.

Таблица 7. Частотности форм дат. и предл. п. ед. ч. во 2 скл.

		Одуш.	Неодуш.	Всего	Одуш.	Неодуш.	Всего
Ж. р.	Дат.	4924	11528	16452	5,9%	3,2%	3,7%
Ж. р.	Предл.	724	32868	33592	0,9%	9,1%	7,5%
Ж. р.	Всего	83322	362027	445349	100,0%	100,0%	100,0%
М. р.	Дат.	1249	—	—	5,7%	—	—
М. р.	Предл.	212	—	—	1,0%	—	—
М. р.	Всего	21940	—	—	100,0%	—	—
Ж. и м. р.	Дат.	6173	11528	17701	5,9%	3,2%	3,8%
Ж. и м. р.	Предл.	936	32868	33804	0,9%	9,1%	7,2%
Ж. и м. р.	Всего	105262	362027	467289	100,0%	100,0%	100,0%

Мы также можем обратить внимание на то, что окончание *-е* используется в дательном падеже единственного числа только во втором склонении, а в предложном падеже единственного числа — также в первом. Общее распределение

этого окончания между различными падежами и числами показано в Таблице 8. А в Таблице 9 мы, наоборот, можем посмотреть, насколько это окончание характерно для того или иного падежа и числа по сравнению с другими окончаниями. База данных позволяет получить эту информацию и с учетом других факторов, например, рода.

Таблица 8. Распределение окончания *-е* между различными падежами и числами

Им. п. ед. ч.	**Дат. п. ед. ч.**	**Вин. п. ед. ч.**	**Предл. п. ед. ч.**	**Им. п. мн. ч.**	**Всего**
15,3%	9,7%	13,4%	61,1%	0,4%	100,0%

Таблица 9. Распределение форм в разных падежах и числах, имеющих окончание *-е*

Им. п. ед. ч.	**Дат. п. ед. ч.**	**Вин. п. ед. ч.**	**Предл. п. ед. ч.**	**Им. п. мн. ч.**	**Всего**
6,2%	26,2%	8,8%	81,4%	0,8%	100,0%

Как нам кажется, приведенные выше примеры показывают, что важнейшее преимущество нашей базы данных в том, что она не только позволяет исследователю ответить на уже поставленные вопросы, но за счет разнообразных способов систематизации информации наталкивает нас на новые вопросы, заставляет рассмотреть новые факторы и задуматься над новыми объяснениями.

5. Заключение

В данной работе представлена база данных, содержащая информацию о частотности форм русских существительных в зависимости от различных грамматических категорий (рода, числа, падежа, одушевленности), от словоизменительных парадигм (склонения и типа основы) и от окончаний. Разные характеристики рассматриваются как по отдельности, так и в различных комбинациях. Предварительная версия базы доступна в интернете (http://www.slioussar.ru/freqdatabase.html). В статье обсуждаются основные принципы сбора информации при создании базы, ряд проблем, с которыми мы столкнулись, а также иллюстрируются некоторые возможности использования содержащейся в базе информации.

Литература

1. *Зализняк А. А.* Грамматический словарь русского языка: Словоизменение. М., 1978.

2. *Копотев М. В.* К построению частотной грамматики: русские падежи по корпусным данным // Инструментарий русистики: корпусные подхода. Хельсинки. 2008. С. 136–151.
3. Русская грамматика. Под ред. Н. Ю. Шведовой. М., 1980.

References

1. *Alegre M., Gordon P.* (1999), Frequency effects and the representational status of regular inflections, Journal of Memory and Language, Vol. 40, pp. 41–61.
2. *Baayen R. H.* (2003), Probabilistic approaches to morphology, in Probability theory in linguistics, MIT Press, Cambridge, MA, pp. 229–287.
3. *Baayen R. H., Dijkstra T., Schreuder R.* (1997), Singulars and plurals in Dutch: Evidence for a parallel dual route model, Journal of Memory and Language, Vol. 36, pp. 94–117.
4. *Baerman M., Brown D., Corbett G. G.* (2005), The syntax-morphology interface: A study of syncretism, Cambridge University Press, Cambridge.
5. *Bybee J.* (2006), Frequency of use and the organization of language, Oxford University Press, Oxford.
6. *Clahsen H.* (1999), Lexical entries and rules of language: A multidisciplinary study of German inflection, Behavioral and Brain Sciences, Vol. 22, pp. 991–1060.
7. *Dressler W. U.* (1985), Morphonology, Karoma Press, Ann Arbor.
8. *Kopotev, M.* (2008), Towards the frequency grammar of Russian: corpus evidence on the grammatical case system [K postroeniyu chastotnoy grammatiki russkogo yazyka: padezhnaya sistema po korpusnym dannym], Instruments of Russian linguistics: corpus approaches [Instrumentariy rusistiki: korpusnye podkhody], Helsinki, pp. 136–150.
9. *Kostič A.* (1991), Informational approach to processing inflected morphology: Standard data reconsidered, Psychological Research, Vol. 53, pp. 62–70.
10. *Kostič A.* (1995), Informational load constraints on processing inflected morphology, in Morphological aspects of language processing.: Lawrence Erlbaum, New Jersey, pp. 317–344.
11. *Lyashevskaya, O.* (2013), Frequency dictionary of inflectional paradigms: core Russian vocabulary, Working papers by Basic Research Programme, Series HUM "Humanities", Moscow.
12. *MacWhinney B., Leinbach J.* (1991), Implementations are not conceptualizations: Revising the verb learning model, Cognition, Vol. 40, pp. 121–157.
13. *Marslen-Wilson W. D., Tyler L. K.* (1997), Dissociating types of mental computation, Nature, Vol. 387, pp. 592–594.
14. *McClelland J.L., Patterson K.* (2002), Rules or connections in past-tense inflections: What does the evidence rule out? Trends in Cognitive Sciences, Vol. 6, pp. 465–472.
15. *Milin P., Filipovic Durdjevic D., Moscoso del Prado Martín F.* (2009), The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian, Journal of Memory and Language, Vol. 60, pp. 50–64.

16. *Moscoso del Prado Martín F., Kostic A., Baayen R. H.* (2004), Putting the bits together: an information theoretical perspective on morphological processing, Cognition, Vol. 94, pp. 1–18.
17. *Pinker S.* (1991), Rules of language, Science, Vol. 253, pp. 530–535.
18. *Pinker S.* (1999), Words and rules: The ingredients of language, Harper Collins, New York.
19. *Pinker S., Prince A.* (1988), On language and connectionism: Analysis of a parallel distributed processing model of language acquisition, Cognition, Vol. 28, pp. 73–193.
20. *Plunkett K., Marchman V.* (1993), From rote learning to system building: Acquiring verb morphology in children and connectionist nets, Cognition, Vol. 48, pp. 21–69.
21. *Rumelhart D., McClelland J. L.* (1986), On learning the past tenses of English verbs, in Parallel distributed processing: Explorations in the microstructure of cognition, MIT Press, Cambridge, MA, pp. 216–271.
22. *Shvedova, N., ed.* (1980), Russian grammar [Russkaya grammatika], Nauka, Moscow.
23. *Ullman M. T.* (2004), Contributions of memory circuits to language: The declarative/procedural model, Cognition, Vol. 92, pp. 231–270.
24. *Zaliznyak, A.* (1978), Grammatical dictionary of Russian language [Grammaticheskij slovar' russkogo yazyka], Moscow.