
КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА И ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА

УДК 004.65+004.438.5

ФОРМАЛЬНЫЙ АНАЛИЗ ПРОЦЕССА ПЛЮРАЛИЗАЦИИ СУЩЕСТВИТЕЛЬНЫХ ПО ДАННЫМ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА

А. А. Кретов

Воронежский государственный университет

Поступила в редакцию 10.02.2018 г.

Аннотация. В статье рассматривается процесс плюрализации существительных по данным Национального корпуса русского языка. Устанавливается зависимость между частотой падежных форм множественного числа и семантикой существительных. Выявляются устойчивые словосочетания и конструкции, влияющие на частоту падежных форм. Аргументируется необходимость их исключения из дальнейшего рассмотрения в целях улучшения результатов семантической классификации слов.

Ключевые слова: компьютерный анализ текста, русский язык, относительная частота падежных форм существительного, формальная семантизация существительных.

Annotation. The article discusses the process of pluralization of nouns according to the data of the Russian National Corpus. The relationship between the frequency of the plural case forms and the semantics of nouns is established. Sustained phrases and constructions that affect the frequency of case forms are identified. The necessity of their exclusion from further consideration in order to improve the results of semantic classification of words is argued.

Keywords: computer analysis of the text, russian language, the relative frequency of case forms of the noun, formal semantization of nouns.

ВВЕДЕНИЕ

Данное исследование имеет прикладное значение для разработки систем автоматической обработки текстов, информационного поиска с поддержкой семантики, систем извлечения знаний из текстов на естественном языке, решения задач снятия семантической неоднозначности.

Семантизация словоформ, несомненно, актуальна, на что указывает присутствие семантической разметки в Национальном корпусе русского языка (далее сокращённо – НКРЯ).

Цель исследования состоит в том, чтобы проверить возможности применения теории, изложенной в работах В. Г. Руделёва [2] и А. Л. Шарандина [3], к материалу Национального корпуса русского языка (далее сокращённо – Корпус).

АНАЛИЗ РЕЗУЛЬТАТОВ ПРЕДШЕСТВУЮЩИХ РАБОТ

Один из вариантов семантической разметки разработан для НКРЯ усилиями Г. Н. Кустовой, О. Н. Ляшевской [14], Е. В. Падучевой и Е. А. Рахилиной [1].

Альтернативная технология семантического текста анализа предложена Ю. Д. Апресяном и его единомышленниками [13]. При этом морфологическая информация, содержащаяся в тексте, не в полной мере используется для семантизации текстовых слов.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Данное исследование выполнено на материале НКРЯ. Объектом исследования являются 11 наиболее употребительных существительных в НКРЯ, у которых самой

употребительной формой является одна из форм множественного числа, **предметом** – относительная частота их словоформ (при абсолютной частоте слов от 1164 до 107).

Метод исследования основан на идее В. Г. Руделёва [2] и А. Л. Шарандина [3] о шифрующей роли морфологии по отношению к лексической семантике, а также – на динамической теории частей речи В. Г. Руделёва [4].

Возмущающим фактором в применении метода является многозначность, чем можно пренебречь, поскольку самое употребительное значение у слова всегда одно, и оно обычно в разы превосходит по частоте ближайшее к нему значение [5].

Гипотеза исследования состоит в предположении, что относительная частота словоформ в парадигме существительного может выступать в качестве классификатора слов по их лексической семантике. При этом словоформы выполняют классифицирующую функцию последовательно: самая употребительная словоформа указывает на класс, менее употребительная по подкласс, третья по употребительности – на подкласс подкласса и т. д.

Функциональная неоднозначность некоторых словоформ является возмущающим фактором, нейтрализуемым, впрочем, у слов одного склонения.

Словоформу, имеющую в парадигме максимальную относительную частоту условимся называть *доминантной* словоформой (ДСлФ), словоформу, занимающую второе место по частоте – *вице-доминантной* (ВДСлФ).

Классификация по ДСлФ даёт первую опозицию: у большинства слов (из 100 самых частотных в НКРЯ) наибольшую частоту имеют словоформы единственного числа, но у 11 слов доминантной является словоформа множественного числа.

Если слова, употребляющиеся исключительно в формах множественного числа, традиционно относят ко множеству *Pluralia tantum*, то слова, в парадигме которых СлФормы множественного числа употребляют-

ся чаще, чем СлФормы единственного условимся называть **плюралитивами**.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

По формам падежа 11 выявленных **плюралитивов** подразделяются на следующие подмножества: 1) Именительный и Именительный/Винительный мн.: *глаз, нога, рука, деньги, результат*; 2) Родительный и Родительный/Винительный мн.: *ребенок, пора, средство, минута*; 3) Творительный мн.: *плечо*; 4) Предложный мн.: *условие*.

Рассмотрим эти множества, начиная со множества, организованного ДСлФ **И./(В.) мн:** *деньги, глаз, нога, рука, результат*.

Одно слово относится к *Pluralia tantum* – *дѣньги* ‘Металлические и бумажные знаки, являющиеся мерой стоимости при купле-продаже’ [6], [9]. Акцентуационно обособленные словоформы единственного числа – *дѣньга, дѣньги, дѣньгой* – лексикализировались и употребляются в значении ‘Старинная ... монета достоинством в полкопейки || Монета любого достоинства.’ [7]. Употребления типа *зашибать/копить деньги* представляют собой традиционные устойчивые сочетания и являются не столько доказательством, сколько пережитком (остатком) значения ‘знаки, являющиеся мерой стоимости при купле-продаже’ у форм единственного числа.

Таким образом, слово *деньги* относятся в *Pluralia tantum* и должно быть исключено из множества **плюралитивов**, коль скоро *деньги* и *деньга* объявляются разными словами – на основании распределения форм единственного и множественного числа между двумя значениями.

Три слова: *глаза, руки, ноги* обозначают **парные части тела**, как правило, функционирующие одновременно, и в силу этого употребляющиеся преимущественно во множественном числе (с тех пор, как утрачено двойственное).

К этим словам примыкают и *плечи*, которых тоже два, правда, у этого слова неожиданно самой употребительной оказывается не форма И-Вмн, а форма Тмн.

ГрЗн	СлФормы	Абс.	Отн.
Тмн	плечами	694	0,24022
И-Вмн	плечи	584	0,20215
И-Вед	плечо	546	0,18899
Пмн	плечах	231	0,07996
Дед	плечу	181	0,06265
Ред	плеча	180	0,06231
Тед	плечом	162	0,05607
Рмн	плеч	135	0,04673
Пед	плече	119	0,04119
Дмн	плечам	53	0,01835
Рмн2	плечей	4	0,00138
	Общий итог	2889	1

Попробуем разобраться, в чём тут дело.

Для этого проанализируем 694 контекста со СлФ плечами:

№	Сочетание	Употр.	Отн.
1	<i>пожать/пожимать (и т. п.) плечами</i>	505	73 %
2	<i>за плечами (у кого-л. что-л.)</i>	131	19 %
3	проч.	58	8 %
	Всего:	694	100 %

«**Пожать** плечами, выразить недоумение, нерешительность, неудовольствие» [8]. «**Пожать плечами, плечом** (выразить недоумение, раздумье, пренебрежение, удивление) [9]».

«За плечами (быть, стоять, иметь и т. п.) 1) В прошлом. 2) В непосредственной близости сзади, позади [9].

Толковые словари дают выражение **пожать/пожимать плечами** при глаголе, а не при существительном, что вряд ли правильно, поскольку, с одной стороны, обедняет фразеологию существительного *плечо*, а с другой, «размазывает» один и тот же фразеологизм по разным словарным статьям.

НКРЯ свидетельствует, что, кроме **пожать/пожимать плечами** имеется немало синонимичных конструкций с той же семантикой и символикой: 'незнания, недоумения, удивления, безразличия'.

Обалдуй слабо пискнул, замялся, глянул куда-то в потолок, **повёл плечами** и умолк. [И. С. Тургенев. Певцы (1850)].

«— Дзержинского, восемь. Водитель недовольно **шевельнул плечами**: — Пешком могли дойти. [Сергей Довлатов. Заповедник (1983)]».

«Ты помнишь, где мы познакомились? Он **подернул плечами**: — В церкви? [Сергей Шаргунов. Оборотень (2011)]».

— И Лора **передёрнула плечами** с выражением мгновенного страха и отвращения. [Юрий Трифонов. Обмен (1969)]

Ее стали утешать, но она **дергала плечами**, а потом начала сметать руками все со своего столика. [Алексей Слаповский. 100 лет спустя. Письма нерожденному сыну // «Волга», 2009] – НКРЯ.

Интересна? Серафима Петровна удивлённо **вскинула плечами**. — С чего вы взяли, что я знаю об этом? [А. Т. Аверченко. Ложь (1910-1911)]

Мы возим. Парень **подвигал плечами**. Молчал. [Роман Сенчин. Елтышевы (2008) // «Дружба Народов», 2009].

В данном случае словоформа *плечами* функционирует не в составе парадигмы слова *плечо*, а в составе конструкций *пожать/пожимать плечами* и (*быть/иметь*) *за плечами* (у кого-л.). Частотность этих словоформ определяется уже семантикой и символикой конструкций, в которых они функционируют. А это уже выводит словоформу *плечами* из системы слова *плечо* и делает её сопоставимой с парадигмой слова *плечо* в целом. Так, *пожать/пожимать плечами*, синонимично глаголам *не знать, удивляться, недоумевать*. А конструкция *за плечами* синонимична речениям *позади, в прошлом*. Как видим, от семантики 'плеча' тут уже практически ничего не остаётся.

Если мы снимем эти инородные напластования, то *плечи* поведут себя, как парный орган человеческого тела, совсем как *глаза, руки и ноги*: *плечи* встречаются чаще, чем *плечо, плечах* – чаще, чем *плече*. И только *плечу* (181) встречается чаще, чем *плечам* (53). И тут причина сходная: на 181 употребление словоформы приходится 140 в конструкции *по плечу*

(что-л. кому-л.): «По плечу кому. Соответствует силам, возможностям, доступно для выполнения» [9]. Эта дефиниция не точна, поскольку описывает не конструкцию *по плечу*, а предикативную конструкцию *быть по плечу кому-л. (о чём-л.)*. Удачнее эта конструкция описана в другом словаре: «ПО ПЛЕЧУ *Предик.* 1. О способности совершить, сделать что-либо благодаря наличию физических сил; в состоянии, в силах, под силу, по силам. 2. *Перен.* О способности совершить, сделать что-либо благодаря наличию умственных, душевных и т.п. возможностей» [7].

В. И. Даль указывает синонимичные конструкции и словоформы для адвербиализованной конструкции: «ПО ПЛЕЧУ *нареч.* под силу, по силе, вмочь; впору, кстати, прилично» [8].

Если же снять и это инородное вкрапление, в Дательном падеже также множественное число (53 употребления) будет преобладать над единственным (41 употребление).

Из сделанных наблюдений следует **алгоритмический вывод**: семантизации существительного по относительной частоте его словоформ должен предшествовать идиоматико-грамматический фильтр – из последующего рассмотрения должны быть исключены употребления слова в составе устойчивых словосочетаний, функционирующих как аналог полнозначного или служебного (грамматического) слова.

Таким образом, одно из множеств **плюралитивов** (с преобладающим Тв мн.) элиминируется, и слово *плечо* включается в первое множество – множество парных предметов, которые можно назвать **дуалитивами** (от лат. *dualis* ‘двойной’). Их типовая семантика – ‘конкретная парность, симметричность, сложность’. Дуалитивы, как и часть слов *pluralia tantum*, называют конкретные предметы (как правило, состоящие из двух симметричных частей): парные части тела (не обязательно, человеческого – *рога, бивни, клыки* и т. д.) Очевидно, что плюралитивы – переходная стадия между «нормальными» словами (с преобладанием СлФорм единственного числа) и словами *pluralia tantum*, у которых формы единственного числа отсутствуют.

Аналогично элиминируется и четвертое множество плюралитивов, состоящее из слова **условие** с преобладающим П. мн.

Это очень интересное слово. БАС-1 выделяет у него 8 значений. Из них второе значение «Официальный договор о чем-л.» имеет помету *Устар.*, три значения имеют помету «*Обычно мн.*» и два значения – «*Чаще мн.*». Таким образом, только два значения – «Уговор, соглашение (устное или письменное) о чем-либо между двумя или несколькими лицами» и «Совокупность данных, положения, лежащие в основе чего-либо» не имеют помет о дефектности парадигмы. Четвертое значение «Обстановка, в которой протекает что-либо, обстоятельства, при которых совершается что-либо» имеет выделенное ромбом употребление «В каких-либо; при каких-либо условиях, в условиях чего-либо и (*устар.*) под какими-либо условиями (делать, совершать и т. п. что-либо или что-либо происходит, протекает и т. п.)» [10].

В более современном словаре [7] конструкция *в условиях чего* уже справедливо трактуется как предлог: «**В условиях** *предл.* с *род.* Соответствует по значению словам: при наличии чего-либо окружающего, обуславливающего или сопутствующего».

Если из 1017 случаев употребления словоформы **условиях** вычесть 859 случаев её употребления с предлогом **в**, остаточная частота этой словоформы (158) уступит не только И-В мн, но и Р мн – *условий*. См. итоговую частотную парадигму слова *условие*.

Падеж	СлФормы	Частота	Отн.
И-В мн	Условия	584	0,39353
Р мн	Условий	298	0,20081
П мн	условиях	158	0,10647
П ед	Услови	119	0,08019
И-В ед	Условие	98	0,06604
Д мн	условиям	94	0,06334
Т мн	условиями	65	0,04380
Тв ед	условием	58	0,03908
Д ед	Условию	10	0,00674
ВСЕГО:	Условие	1484	1,00000

Как видим, повышенная употребительность словоформы условиях вызвано её грамматикализацией и употреблением в составе предлога **в условиях** чего-л. или **в** (каких-л.) **условиях**.

Таким образом, от 4 множеств плюралитивов, у нас остаётся 2.

Последнее слово первого множества плюралитивов с доминантой И-В. мн. – **результаты** – требует отдельного разговора.

Преимущественная сфера его функционирования – наука, техника, производство: 69 % словоупотреблений в НКРЯ.

Тематически также преобладает наука и виды точного знания: 58 %.

Но самым выразительным показателем является тип текста: *статья* (47 %) и *монография* (8 %), что в сумме сразу даёт 55 %. Добавление к этому типов: *учебное пособие, отчет, учебник, рецензия, доклад, лекция | цикл, монография | статья* увеличивает долю научных типов текста до 60 %.

Научное исследование строится на принципе воспроизводимости *результатов*, поэтому один *результат* в науке, строго говоря, *результатом* не является. Доказательна только представительная серия *результатов*. В этом, видимо, и состоит причина преобладания множественного числа этого слова над единственным. При этом *результаты* мыслятся (и подаются) как некая самостоятельная сущность или самостоятельная тема высказывания, способная занимать в предложении позицию подлежащего. Отсюда – преобладание именительного падежа. Чаще всего речь идёт о *результатах анализа/анализов, выборов, голосования, деятельности, измерения/*

измерений, изучения, испытания/испытаний, исследования/исследований, наблюдения/наблюдений, обследования/обследований, опроса/опросов, опытов, работы/работ, расчета/расчетов, решения/решений, эксперимента/экспериментов.

Если специализация слова усилится, наметившийся процесс «плюрализации парадигмы» может привести к семантической дифференциации парадигмы и превращению слова результаты в *Pluralia tantum*.

Проанализируем парадигму слова результат (табл. 1).

Как видим, в четырёх падежах из пяти преобладает множественное число, и только в творительном и предложном преобладает единственное число. При этом, если Тв. ед. встречается в 2 раза чаще чем Тв. мн., то П. ед. – более, чем в 20 раз!

Причиной этого является формирование сложных служебных слов (союзов и предлога) на базе предложно-падежной конструкции *в результате*:

«В результате, в знач. союза. В итоге. Родион Антонович тоже стрелял, и его Зарез [собака] работал на славу; в результате оказалось, что он убил [дупелей] больше всех. Мам.-Сиб. Горн. Гнездо.

В результате того, что..., в знач. союза. Опоздали в результате того, что не рассчитали время.

В результате чего-л., в знач. предлога с род. пад. Вследствие чего-л, благодаря чему-л. <...> Человек в результате сотрясения мозга может забыть родной язык. К. Платонов, Занимат. психология». [БАС т. 23:572-573].

Таблица 1

Суммарное представление частот словоформ слова **результат** (по данным 10 выборок из НКРЯ)

Падеж	Ед. число	СлУпЕд	Мн. число	СлУпМн	Ед/Мн
И/В	результат	475	результаты	541	0,88
Р	результата	129	результатов	221	0,58
Д	результату	28	результатам	114	0,25
Т	результатом	118	результатами	56	2,11
П	результате	824	результатах	40	20,60
Всего:		1574		972	

Чтобы получить собственно лексическую закономерность, влияние грамматических явлений, к которым относятся предлоги и союзы, надо снять. Без форм П. падежа мы получаем соотношение 750 (ед.) – 932 (мн.). Формы множественного числа преобладают над формами единственного в 1,2 раза. Формы множественного числа преобладают по меньшей мере в 3 случаях из 5. На этом основании можно сделать вывод, что наметившийся процесс «плюрализации парадигмы» и превращения слова *результат* в *Pluralia tantum* ближе к концу, что к началу.

Для получения более надёжной информации, мы проанализировали 10 выборок словоупотреблений слова *результат* в НКРЯ, сведения о которых представлены в табл. 2.

Рассмотрим борьбу за употребительность между словоформами слова *результат* по падежам.

На рис. 1. представлено соотношение форм ед. и мн. числа Именительного/Винительного падежей.

Как свидетельствует рис. 1, лишь в трёх выборках из 10 форма ед. ч. употребляется чаще, чем форма множественного числа.

Таблица 2

Раздельное представление частот словоформ слова *результат*
(по данным 10 выборок из НКРЯ)

№	Выборка СлФ	абсолютн.									
		1	2	3	4	5	6	7	8	9	10
1	результат	54	35	44	65	46	64	55	34	52	26
2	результата	13	5	10	13	8	10	11	8	6	0
3	результатам	21	14	8	7	9	13	15	6	18	3
4	результатами	10	8	7	4	2	4	4	3	6	8
5	результатах	3	5	6	2	4	8	5	1	3	3
6	результате	98	81	70	70	120	76	76	62	97	74
7	результатов	44	31	23	15	15	29	11	14	22	17
8	результатом	19	13	13	12	13	16	6	5	13	8
9	результату	8	1	3	4	1	4	1	1	3	2
10	результаты	103	55	68	37	49	59	23	43	60	44
Σ	результат	373	248	252	229	267	283	206	177	280	185

№	Выборка СлФ	абсолютн.									
		1	2	3	4	5	6	7	8	9	10
1	результат	0,14	0,14	0,17	0,28	0,17	0,23	0,27	0,19	0,19	0,14
2	результата	0,03	0,02	0,04	0,06	0,03	0,04	0,05	0,05	0,02	0,00
3	результатам	0,06	0,06	0,03	0,03	0,03	0,05	0,07	0,03	0,06	0,02
4	результатами	0,03	0,03	0,03	0,02	0,01	0,01	0,02	0,02	0,02	0,04
5	результатах	0,01	0,02	0,02	0,01	0,01	0,03	0,02	0,01	0,01	0,02
6	результате	0,26	0,33	0,28	0,31	0,45	0,27	0,37	0,35	0,35	0,40
7	результатов	0,12	0,13	0,09	0,07	0,06	0,10	0,05	0,08	0,08	0,09
8	результатом	0,05	0,05	0,05	0,05	0,05	0,06	0,03	0,03	0,05	0,04
9	результату	0,02	0,00	0,01	0,02	0,00	0,01	0,00	0,01	0,01	0,01
10	результаты	0,28	0,22	0,27	0,16	0,18	0,21	0,11	0,24	0,21	0,24
Σ	результат	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

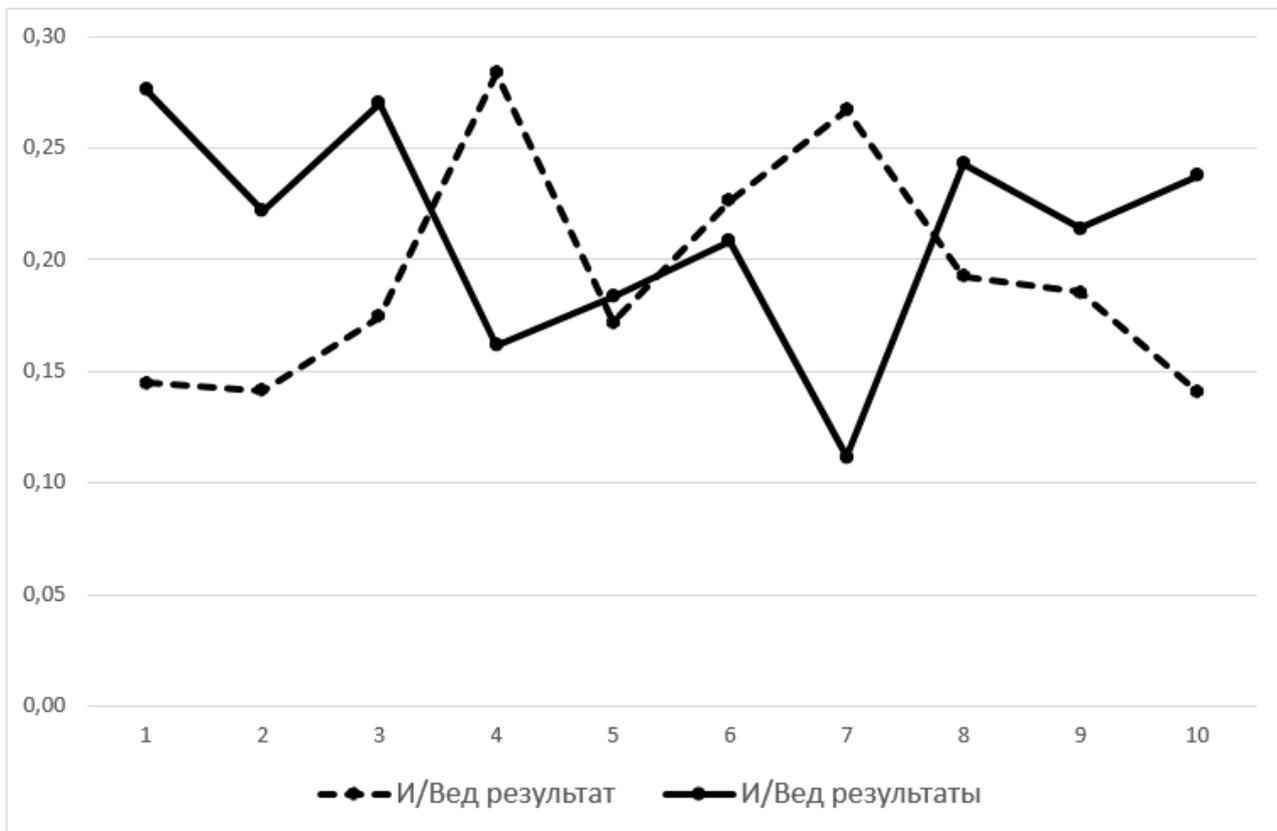


Рис. 1. Соотношение форм ед. и мн. числа Именительного/Винительного падежей (по данным 10 выборок из НКРЯ)

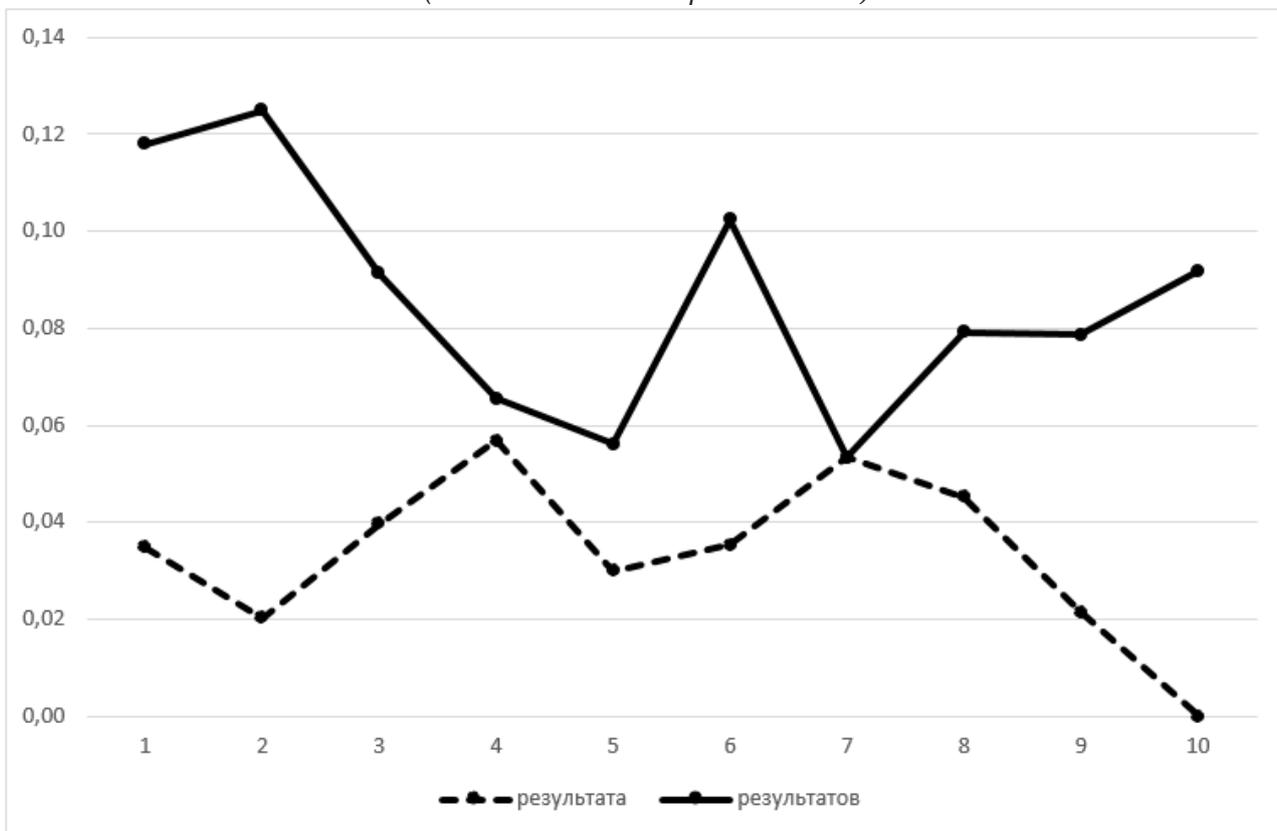


Рис. 2. Соотношение форм ед. и мн. числа Родительного падежа (по данным 10 выборок из НКРЯ)

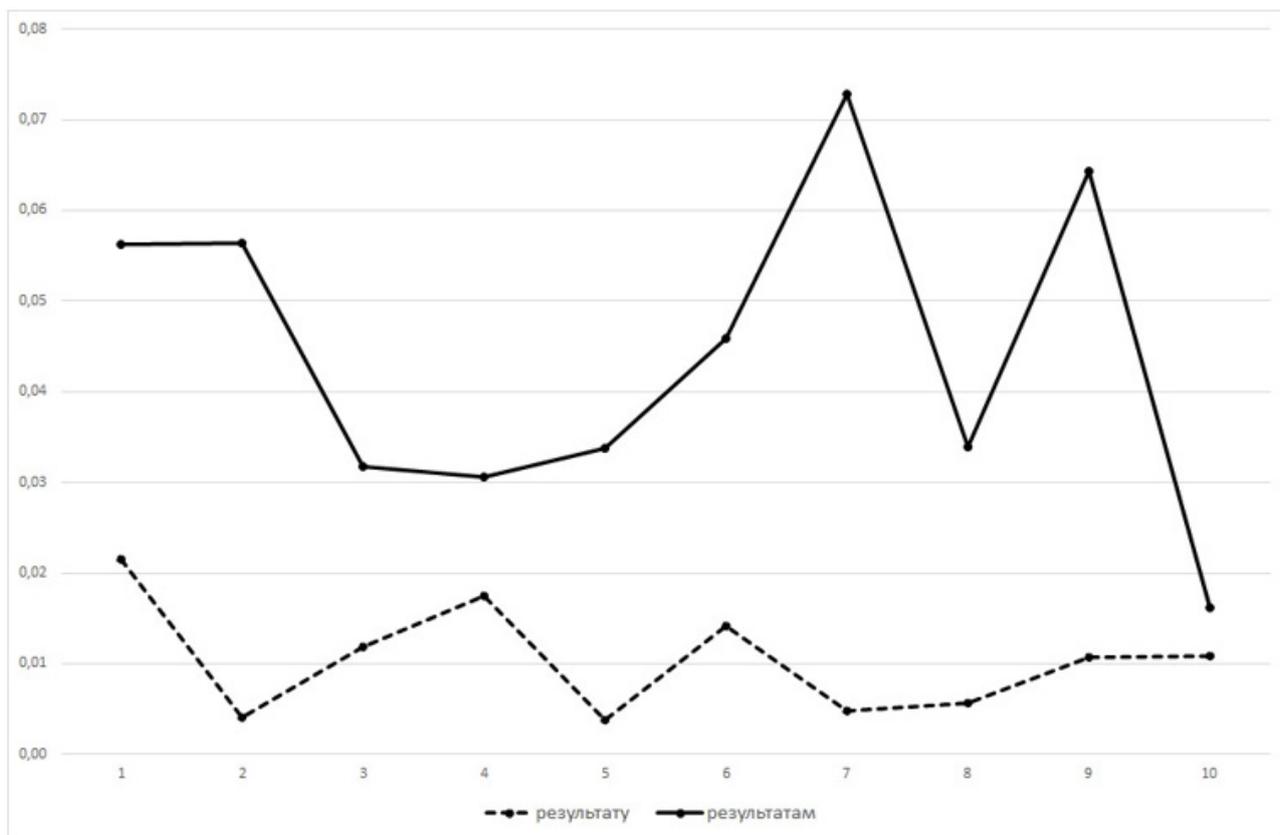


Рис. 3. Соотношение форм ед. и мн. числа Дательного падежа (по данным 10 выборок из НКРЯ)

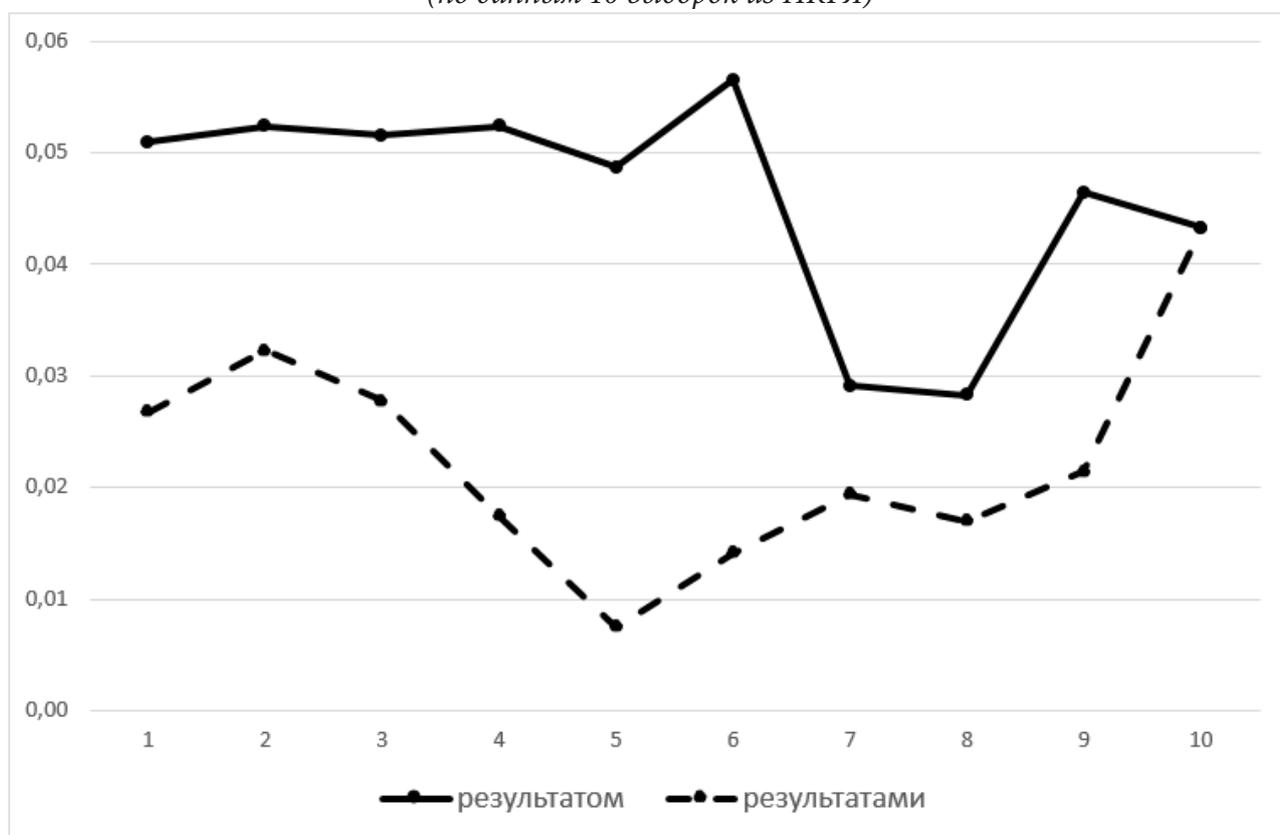


Рис. 4. Соотношение форм ед. и мн. числа Творительного падежа (по данным 10 выборок из НКРЯ)

Борьба за употребительность близка к завершению в пользу формы множественного числа. Причина трёх отклонений от общей закономерности может состоять в том, что ед. число преобладает тогда, когда СлФ чаще употребляется в значении В.п.; в функции И.п. вероятнее преобладание формы мн.ч.

О том, что мы имеем дело именно с борьбой словоформ, свидетельствует отрицательная корреляция между формами: возрастание употребительности одной из форм сопровождается понижением употребительности другой. Пожалуй, столько в выборках 1, 2, 6, 9 эта корреляция не просматривается.

На рис. 2 представлено соотношение форм ед. и мн. числа Родительного падежа.

Как видим, в Родительном падеже победа множественного числа над единственным уже состоялась. Лишь в одной выборке относительная частота СлФ равна, в остальных преобладает множественное число. При этом отрицательная корреляция СлФ очевидна. Лишь в выборках 5, 6, 9 она не просматривается.

На рис. 3 представлено соотношение форм ед. и мн. числа Дат. падежа.

Борьба за дательный падеж завершена полностью и окончательно в пользу множественного числа. Корреляция между формами уже почти (за исключением выборок 3, 5, 7) не просматривается.

Соотношение форм ед. и мн. числа Творительного падежа представлено на рис. 4.

Борьба за творительный падеж (кстати, говоря, самый устойчивый в системе русского склонения) находится в самом начале. Об этом свидетельствует не контрастная, а параллельная корреляция относительных частот в выборках: повышение и понижение употребительности происходит одновременно в формах Ед и мн. числа. Лишь в двух случаях наблюдается контрастная корреляция: выборки 7 и 10. В Выборке 10 относительные частоты падежей уже оказываются равны. Творительный падеж позволяет наблюдать самое начало процесса плюрализации.

Соотношение форм ед. и мн. числа Предложного падежа представлено на рис. 5.

В предложном падеже за явным преимуществом побеждает единственное число. Корреляция между употребительностью форм практически отсутствует. Правда, причины этого явления в грамматикализации (употреблении в значении союзов и предлога) предложно-падежной конструкции в *результате*.

Для обеспечения сопоставимости результатов необходимо было бы исключить из рассмотрения все случаи союзного и предложного употребления СлФ в *результате*. В данном случае в этом нет необходимости так как общее преобладание форм множественного числа и направление процесса плюрализации слова *результат* налицо.

Схематически процесс *плюрализации существительного* выглядит так. Сначала формы ед. и мн. числа ведут себя сходным образом: их частота зависит от употребительности других падежей.

При этом имеет место параллельная корреляция. Свидетельством начала конкуренции чисел является **появление отрицательной корреляции**, когда частота одного из чисел внутри падежа зависит от частоты другого.

При этом появляется возможность сближения частот чисел в отдельных выборках. Затем от выборки к выборке появляется преобладание множественного числа над единственным. Завершает процесс абсолютным преобладанием форм множественного числа над единственным и утратой корреляции между употребительностью СлФ числа внутри падежа.

Этот процесс сопровождается и завершается (закрепляется) расхождением лексической семантики словоформ единственного и множественного числа. Суммарное соотношение функционирования СлФ слова *результат* представлено на рис. 6.

Во втором множестве (Родительный и Родительный/Винительный мн) у нас 4 слова: *пора, минута, средство, ребенок*.

Слово **пора** из этого множества также следует удалить: все употребления слова в форме Род. мн. (а их 1351) приходятся на фразеосочетания *с тех / этих / каких / давних / недавних / некоторых пор и до сих / тех / каких /*

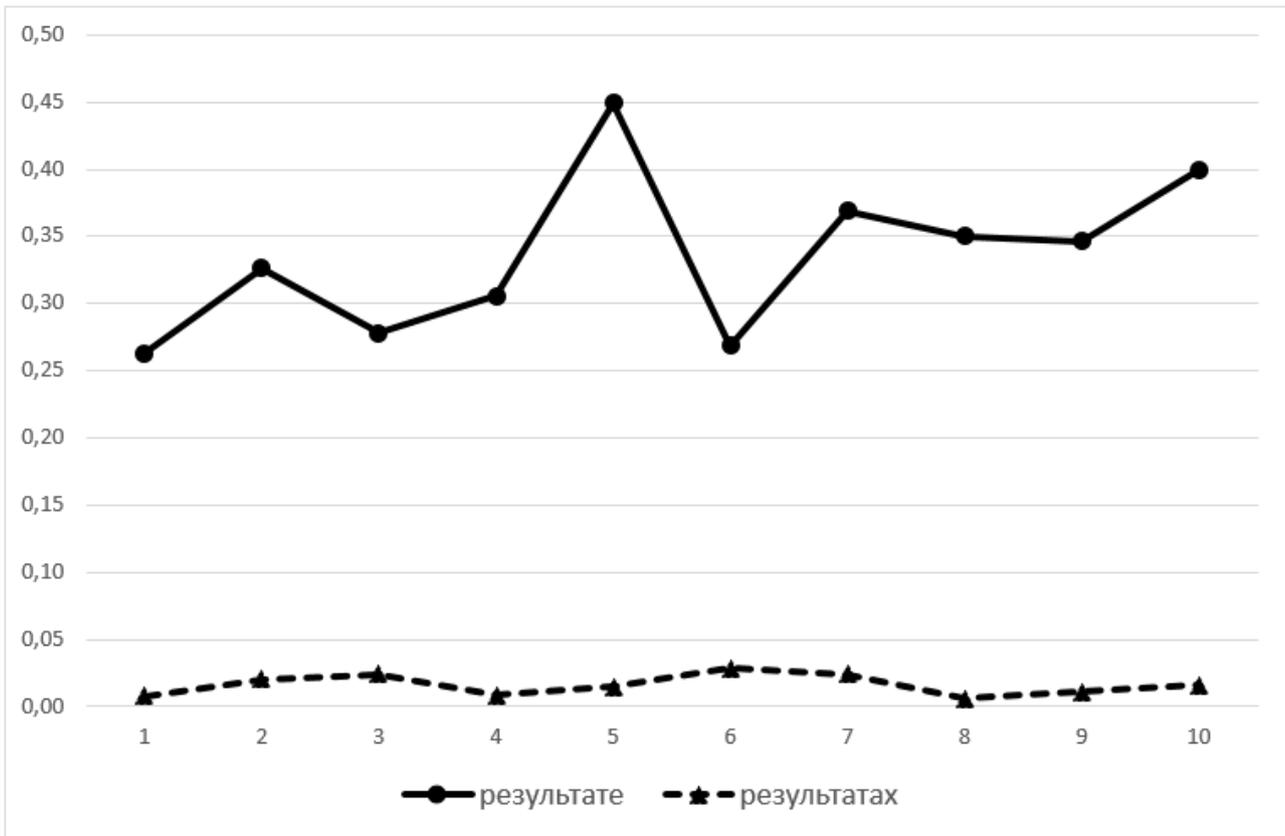


Рис. 5. Соотношение форм ед. и мн. числа Предложного падежа (по данным 10 выборок из НКРЯ)

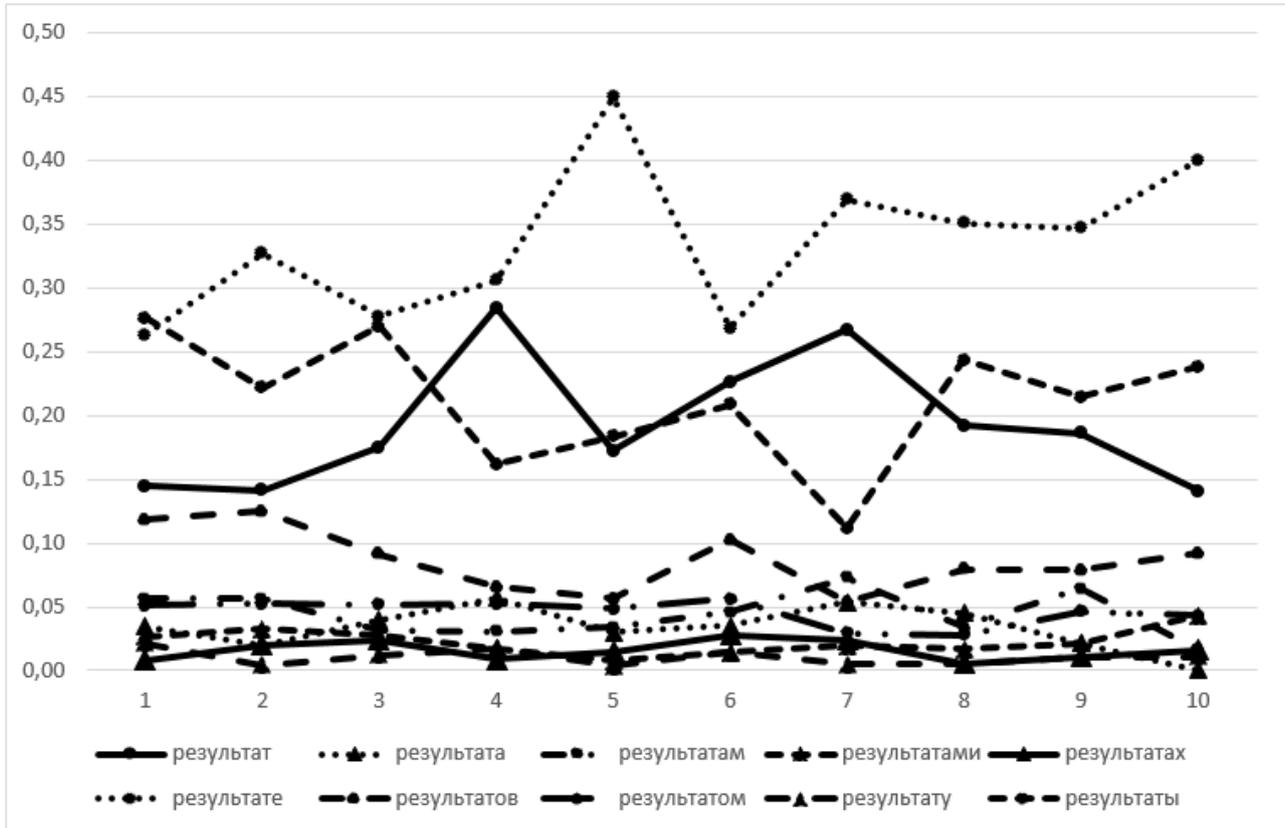


Рис. 6. Соотношение всех СлФ слова результат (по данным 10 выборок из НКРЯ)

недавних пор. Архаизмами являются варианты по сих пор и по тех пор, встретившиеся в НКРЯ по 1 разу.

Речение	Раз
до сих пор	779
с тех пор	342
до тех пор	165
с некоторых пор	20
с этих пор	13
с давних пор	10
с каких пор	8
до недавних пор	6
с недавних пор	5
до каких пор	1
по сих пор	1
по тех пор	1
Общий итог	1351

После их исключения частотная парадигма слова *пора́* выглядит следующим образом:

Падежи	Словоформы	Частота	Относит.
И ед	пора	432	0,38196
Т ед	порой	277	0,24492
В ед	пору	215	0,19010
Р ед	поры	113	0,09991
Т' ед	порою	47	0,04156
П мн	порах	32	0,02829
Д ед	поре	15	0,01326
	пора́	1131	1,00000

Если принять во внимание, что все 32 случая употребления П мн приходится на фразеологизм *на первых порах*, то трудно отделаться от впечатления, что слова *пора́* и *пóры* 'мельчайшие отверстия на поверхности кожи' поделили между собой парадигму: *пора́* взяла себе единственное число, а *пóры* – множественное.

Мнимым плюралитивом оказывается и слово *минута*:

Падежи	Словоформы	Раз
Р мн	минут	1315
В ед	минуту	819
Р ед-И мн	минуты	547

И ед	минута	85
Д ед	минуте	33
П мн	минутах	30
Т ед	минутой	26
Т мн	минутами	15
Д мн	минутам	6
Т' ед	минутую	3
	Общий итог	2879

Всё богатство её употреблений в Р. мн. связано с контекстом числительного: 5-20 и т. д. *минут* или 2-3-4 *минуты*.

После вычета этих конструкций из суммарной частоты слова и словоформ слово *минута* не попадает во множество плюралитивов.

Рассмотрим пару *дети-ребёнок*. Тут всё непросто, поэтому будем пользоваться показателями самого нового и самого авторитетного словаря – БАСа-3 [6].

К словарной форме *ребёнок* БАС-3 даёт в качестве множественного числа *дети* и *ребята*.

А к словарной форме *дети* (в БАСе-3 это словарная форма) – в качестве единственного – *дитя* и *ребёнок*. К словарной форме *дитя* даётся множественное *дети*.

Вряд ли обосновано приведение БАСом-3 (со ссылкой на словарь БТС 1998) словарной формы *дитятя*, пусть даже и с пометой *Прост.*, тем более что в иллюстративном примере приводится словоформа *дитятей* со значением «О наивном, неопытном человеке», предусмотренная парадигмой слова *дитя*. Её уместнее было бы дать как одно из (под) значений слова *дитя*, что в словарной статье *дитя*, кстати говоря, и сделано. А уж если материал картотеки БАСа-3 позволяет выделять такую экзотическую форму И.ед. как *дитятя*, то именно её следовало бы подтвердить иллюстративным материалом в первую очередь, чтобы на читателя эта форма не производила впечатление лексикографического фантома. (НКРЯ даёт 11 случаев употребления этой формы. Из них три – у В. Ф. Одоевского «Сказки дедушки Ириней (1841)», а остальные 8 – у авторов с 1985 по 2014 год. «Например, раздражает постоянно повторяющееся слово «**дитятя**» — откуда оно могло

взяться в современном переводе с французского?» [Олеся Ханцевич. Французский психолог снова рекомендует // «Эксперт», 2014]; «Венька Фомин, как дитятя от родительского ремешка, пытался прикрыть зад ладонями». [Виктор Астафьев. Печальный детектив (1982–1985)].

Ситуацию окончательно запутывает то, что в БАСе-3 в качестве словарной даётся также форма РЕБЯТА с грамматической пометой «мн.» с двумя значениями: 1) мн. ч. к *ребёнок*, 2) Разг. Молодые люди, парни. Помета *мн.* при словарной форме, должна, видимо, указывать на то, что это слово *Pluralia tantum*, но отсылка первого значения к ед. *ребёнок* дезавуирует эту догадку. Видимо, корректнее было бы поставить помету «мн.» при втором значении.

Указанные лексикографические сложности вызваны тем, что у нас на глазах из двух однородных парадигм: *ребёнок-ребята* и *дитя-дети* формируется супплетивная парадигма *ребёнок-дети*. Слово *дитя* (в значительной степени из-за архаичности и нерегулярности своей парадигмы в единственном числе) постепенно выходит из употребления, а *ребята* (вместе с другими формами множественного числа) в значении «Молодые люди, парни» понемногу превращается в слово *Pluralia tantum*.

Если мы проанализируем частотную парадигму слов *ребёнок*, *ребята* и *дитя*, в этом легко убедиться.

ГрамЗнач	СлФормы	Раз	Относит.
Р-Вмн	детей	1120	0,334
Имн	дети	829	0,247
Р-Вед	ребёнка	420	0,125
Иед	ребёнок	300	0,090
Дмн	детям	226	0,067
Тмн	детьми	203	0,061
Тед	ребёнком	112	0,033
Дед	ребёнку	83	0,025
Пмн	детях	46	0,014
Пед	ребёнке	12	0,004
	ребёнок	3351	1,000

В обоих числах преобладают формы Р-В падежей, что свидетельствует о наличии у

слова семантических признаков 'несамостоятельное одушевлённое'.

Падеж	СлФорма	Раз	Относит.
Р-Вмн	<i>детей</i>	1485	0,4484
Имн	<i>дети</i>	1115	0,3367
Дмн	<i>детям</i>	262	0,0791
Тмн	<i>детьми</i>	256	0,0773
И-Вед	<i>дитя</i>	116	0,0350
Пмн	<i>детях</i>	68	0,0205
Ред	<i>дитяти</i>	5	0,0015
Дед	<i>дитяти</i>	4	0,0012
Тед	<i>дитем</i>	1	0,0003
	ДИТЯ	3312	1,0000

Здесь формы множественного числа с преобладанием Р-В падежа также указывают на несамостоятельность, а формы единственного – указывают о преобладании И-В падежа, что обычно свидетельствует о самостоятельности, а данном случае на частоту влияет не семантика, а грамматика: говорящие избегают архаичной формы *дитяти*, и стремятся выравнять парадигму либо по косвенным падежным формам единственного числа (ср. форму И ед. *дитятя*, предлагаемую БАСом-3) либо по множественному числу (ср. форму *дитём*, обнаруживаемую в НКРЯ: «Завтра Таньке напишу, но куда она с малым **дитём** из Воронежа поедет?» [Андрей Волос. Недвижимость (2000) // «Новый Мир», 2001]). По существу, тут реализована морфологически регулярная форма среднего рода *дитё*, которую словари также дают отдельной словарной формой – с пометой *разг.-сниж.* (Об орфографической оправданности написаний *дитя* и *дитё* говорить трудно, потому что ударная форма **дѣти** указывает на *Е*, а значит, по общему правилу при наличии «проверочного слова» следовало бы писать **дет'я* и **детѣ*, как мы пишем *вода* и *водичка*, потому что есть *вѣды*. Появление *И* в корне можно объяснить влиянием украинских или северно-русских говоров, нашедшем отражение в нормах русской орфографии).

Наконец, частотная парадигма слова *ребята* наглядно свидетельствует об отрыве этой парадигмы от семантики несамостоятельного

одушевлённого, поскольку форма И мн. более чем вдвое превосходит по частоте форму Р-В мн.

ГрамЗн	СлФормы	Раз	Относит.
Имн	<i>ребята</i>	1487	0,601
Р-Вмн	<i>ребят</i>	645	0,261
Дмн	<i>ребятам</i>	169	0,068
Тмн	<i>ребятами</i>	166	0,067
Пмн	<i>ребятах</i>	7	0,003
	ребята	2474	1,000

Преобладание И п. свидетельствует о семантике активности и самостоятельности.

Так употребительность словоформ подтверждает правомерность лексикографических решений: плюралитив *ребята* 'Разг. Молодые люди, парни' семантически и функционально разошёлся с сингулятивом *ребёнок* 'Мальчик или девочка в раннем возрасте, до отрочества', которому в связи с этим пришлось задействовать и адаптировать формы множественного числа стремительно устаревающего и выходящего из употребления существительного *дитя*, что уже фактически привело к образованию семантически однородной супплетивной парадигмы *ребёнок-дети*.

Пик употребительности словоформы *дитя* пришёлся на 1830-ые годы (эпоху романтизма). С 1870 (эпоха разночинцев) её употребительность резко пошла на убыль. В 1910-ые годы (Серебряный век) она пережила небольшой всплеск употребительности, который быстро пошёл на убыль и впоследствии

сколь-нибудь существенных всплесков употребительности этой словоформы не наблюдалось.

Во втором множестве (Рмн) осталось рассмотреть слово *средство*.

Падежи	СлФормы	Раз	Относит.
Рмн	<i>средств</i>	954	0,415
Ред/Имн	<i>средства</i>	813	0,354
Иед	<i>средство</i>	185	0,081
Тмн	<i>средствами</i>	184	0,080
Тед	<i>средством</i>	74	0,032
Пмн	<i>средствах</i>	56	0,024
Дмн	<i>средствам</i>	24	0,010
Дед	<i>средству</i>	4	0,002
Пед	<i>средстве</i>	3	0,001
	средство	2297	1,000

Как мы видим, это самый настоящий плюралитив: все падежи чаще употребляются во множественном числе, чем в единственном. Несколько осложняет анализ совпадение форм Ред и Имн. но даже в этом случае Рмн употребляется чаще, чем два этих падежа вместе взятых.

МАС-2 [11] выделяет у слова *средство* 6 значений. Из них 2 («Устар. Способности, данные» и «Деньги, капитал, материальные ценности») употребляются только во множественном числе, а одно – («Предметы, приспособления или совокупность их, необходимые для осуществления чего-л.») «обычно во мн.».



Рис. 7. Динамика словоформы *дитя* по данным НКРЯ

Один из новейших словарей трактует слово *средства* в значении 'Деньги, материальные ценности' уже как *Pluralia tantum*: «СРЕДСТВА, средств, только мн. Деньги, материальные ценности; Син.: капитал» [12].

Анализ данных НКРЯ свидетельствует, что для этого имеются основания: из 954 случаев употребления формы *средств* в НКРЯ на долю этого значения приходится 807 употреблений, т. е. 86 %.

При этом львиная доля употреблений приходится на конструкции *из средств, за счёт средств, без/нет/много средств* и *доля/часть/остаток/сумма средств*.

В данном случае *средства* предстают не как самостоятельный деятель, а как 'ресурс, мыслимый в качестве инактивной субстанции, необходимой для достижения, осуществления чего-либо'.

Таким образом, во множестве ДСлФР(В) мн осталось два слова: *ребенок* и *средство*. Оба они объединены семантикой 'несамостоятельности'.

ЗАКЛЮЧЕНИЕ

1) Исследование частотности субстантивных парадигм позволяет в первом приближении установить семантику слова.

2) Преобладание частотности у форм множественного числа свидетельствует либо о принадлежности слова к множеству парных предметов: *глаз, нога, рука, плечо*. Либо к абстрактным существительным – *результат, средства*.

3) Если преобладание Имн свидетельствует о семантике 'парности' или 'абстрактности' в сочетании с 'активностью и самостоятельностью', то преобладание Рмн (*ребёнок, средства*) свидетельствует о семантике 'пассивности и несамостоятельности'.

4) Семантизации существительного по относительной частоте его словоформ должен предшествовать идиоматико-грамматический фильтр: из последующего рассмотрения должны быть исключены употребления слова в составе устойчивых словосочетаний, функционирующих как аналог полнозначного или служебного (грамматического) сло-

ва. Классификации имени по доминантной словоформе должно предшествовать *снятие фразеологии и грамматики*, состоящее в исключении из подсчётов употребления словоформ в составе более сложных и устойчивых единиц.

5) Определение количественного порога, свидетельствующего о наличии устойчивых сочетаний – отдельная прикладная задача. Пока можно предложить использовать полученные «вручную» результаты анализа рассмотренных выше слов.

6) Словам с полной парадигмой противопоставлены слова с ущербной парадигмой – только единственного числа (*Singularia tantum*) или только единственного числа (*Pluralia tantum*). Нормой является преобладание форм единственного числа над формами множественного. Промежуточное положение между нормальными словами и *Pluralia tantum* занимают слова с преобладанием форм множественного числа при наличии форм единственного числа. Такие слова предложено называть *плюралитивами*, а слова, называющие парные предметы – *дуалитивами*. При этом *плюралитив* – гипероним, а *дуалитив* – гипоним. Процесс постепенного ограничения парадигмы слова формами множественного числа предлагается называть *процессом плюрализации*.

7) Схематически процесс *плюрализации* существительного выглядит так.

Сначала формы ед. и мн. числа ведут себя сходным образом: их частота зависит от употребительности других падежей. При этом имеет место параллельная корреляция падежей в ед. и мн. числа.

Свидетельством начала конкуренции чисел является **появление отрицательной корреляции**, когда частота одного из чисел внутри падежа зависит от частоты другого.

При этом появляется возможность сближения частот чисел в отдельных выборках. Затем от выборки к выборке появляется преобладание множественного числа над единственным. Завершает процесс абсолютным преобладанием форм множественного числа над единственным и утратой корреляции между употребительностью СлФ числа внутри падежа.

Этот процесс сопровождается и завершается (закрепляется) расхождением лексической семантики словоформ единственного числа (сингуляризацией) и множественного числа (плюрализацией).

Исследование выполнено в рамках проекта «Лингвистическая подготовка и создание электронного «Живого стилистического словаря русского языка», поддержанного грантом РФФИ № 17-04-00421.

СПИСОК ЛИТЕРАТУРЫ

1. Кустова Г. Н., Ляшевская О. Н., Падучева Е. В., Рахилина Е. А. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. – М. : Индрик, 2005. – С. 155–174.

2. Руделёв В. Г., Шарандин А. Л. Шифрующая роль глагольных грамматических категорий // Теория содержательной формы. – Тамбов, 1981. – С. 32–54.

3. Шарандин А. Л. Лексико-семантические разряды русского глагола // Семантика языковых единиц. – Т. 2. – М., 1998. – С. 46–50.

4. Руделёв В. Г. Динамическая теория частей речи русского языка // обр. соч. в 6 томах. Т. 6. Друг мой – язык. Избр. произв. по общему и русскому языкознанию. – Тамбов : Изд-во ТГУ им. Г.Р. Державина, 2002. – С. 36–49.

5. Кретов А. А., Терентьева И. А. Распределение функциональной нагрузки между значениями многозначных слов // Проблемы компьютерной лингвистики : Сборник научных трудов / Под ред. А. А. Кретова. – Вып. 4. – Воронеж, 2010. – С. 293–301.

6. Большой академический словарь русского языка. В 30 томах / Гл. ред. К. С. Горба-

чевич, А. С. Герд, С. А. Мызников. – М., СПб. : Наука, 2004.

7. Ефремова Т. Ф. Большой современный толковый словарь русского языка. 180 тыс. статей. © 2006, [Explanatory (Ru-Ru) (к версии Lingvo 12)].

8. Толковый словарь живого великорусского языка : в 4 т. / авт.-сост. В. И. Даль. – 2-е изд. – СПб. : Типография М. О. Вольфа, 1880–1882. Том 3 (1882).

9. Большой толковый словарь русского языка © С.А. Кузнецов, 2010 [ExplanatoryBTS (Ru-Ru) (к версии АBBYY Lingvo x5)].

10. Словарь современного русского литературного языка: В 17 т. / Под ред. В. И. Чернышёва. – М., Л. : Изд-во АН СССР, 1948–1965.

11. Словарь русского языка: В 4-х т. / РАН, Ин-т лингвистич. исследований; Под ред. А. П. Евгеньевой. – 4-е изд., стер. – М. : Рус. яз.; Полиграфресурсы, 1999.

12. Большой толковый словарь русских существительных: Идеографическое описание. Синонимы. Антонимы / Под ред. проф. Л. Г. Бабенко. – М. : АСТ-ПРЕСС КНИГА, 2005. – 864 с.

13. Апресян Ю. Д., Иомдин Л. Л., Санников А. В., Сизов В. Г. Семантическая разметка в глубоко аннотированном корпусе русского языка // Труды Международной конференции «Корпусная лингвистика – 2004. – СПб. : Изд-во Санкт-Петербургского университета, 2004. – С. 41–54.

14. Ляшевская О. Н. Семантика русского числа. – М. : Языки слав. Культуры, 2004. – 362 с.

Кретов Алексей Александрович – д-р филол. наук, проф. каф. теоретической и прикладной лингвистики, Воронежский государственный университет.
Тел.: (+7 4732) 204-149
E-mail: tipl@rgph. vsu.ru

Kretov A. A. – Doctor of Philology Sciences, Professor of the dept. of Theoretical and Applied Linguistics, Voronezh State University.
Tel.: (+7 4732) 204-149
E-mail: tipl@rgph.vsu.ru