

А. Я. ШАЙКЕВИЧ, В. М. АНДРЮЩЕНКО, Н. А. РЕБЕЦКАЯ

**ДИСТРИБУТИВНО-
СТАТИСТИЧЕСКИЙ АНАЛИЗ
ЯЗЫКА РУССКОЙ ПРОЗЫ
1850—1870-х гг.**

Том 1

Studia philologica

УДК 811.161.1
ББК 81.2 Рус
Ш 17

Издание подготовлено при финансовой поддержке
Российского гуманитарного научного фонда (РГНФ)
проект № 13-04-16009

Рецензенты:
д. ф. н. А. Ф. Журавлев, д. ф. н. С. А. Крылов

Шайкевич А. Я., Андриющенко В. М., Ребецкая Н. А.

Ш 17 Дистрибутивно-статистический анализ языка русской прозы 1850—1870-х гг. Т. 1. М.:
Языки славянской культуры, 2013. — 504 с. — (Studia philologica.)

ISBN 978-5-9551-0668-7

Цель дистрибутивно-статистического анализа состоит в открытии структуры языка на основе большого корпуса текстов. В настоящей трехтомной монографии этот формальный метод в полной мере прилагается к текстам русской прозы 1850—1870 гг. (около 15 млн словоупотреблений); а частично (в виде иллюстраций) к текстам на других языках.

Первый том включает три части:

Очерк развития метода;

Открытие регулярной морфологии в рамках графического слова;

Частотный словарь языка русской прозы 1850—1870 гг.

Первые две части адресованы лингвистам, особенно тем, кто интересуется лингвостатистикой. Частотный словарь будет интересен филологам-русистам. В существенно расширенном виде он представлен на компакт-диске.

ББК 81.2 Рус

Электронная версия данного издания является собственностью издательства,
и ее распространение без согласия издательства запрещается.

ISBN 978-5-9551-0668-7

© А. Я. Шайкевич, В. М. Андриющенко, Н. А. Ребецкая, 2013
© Языки славянской культуры, 2013

Часть 2

**ДИСТРИБУТИВНО-СТАТИСТИЧЕСКИЙ АНАЛИЗ
В МИКРОИНТЕРВАЛЕ
(СТАТИСТИЧЕСКОЕ ОТКРЫТИЕ РЕГУЛЯРНОЙ МОРФОЛОГИИ)**

В первой части книги (особенно с. 32, 34, 42, 53—56) были намечены пути формального открытия морфологии в пределах графического слова. Можно предложить множество подобных путей анализа, опирающихся на статистику комбинаторики букв в буквенных цепочках и на вероятность появления фрагментов таких цепочек в определенных окружениях.

В качестве первого примера рассмотрим подход, учитывающий статистику отдельных букв и буквосочетаний в их связи с началом и концом графического слова¹. Соответствующим корпусом послужил комплект «Независимой газеты» за 1996—2000 гг. (35 млн словоупотреблений). Из общего частотного словаря выделена зона частых слов с вероятностью ниже 0,001 и выше 0,000001 (всего более 62 тысяч разных графических слов).

2.1. Позиции, n-граммы и парадигмы в статистическом открытии морфологии

В общем случае исследователь в рамках ДСА опирается на буквы (с их частотами), на какие-то заданные позиции в тексте (например, рядом с пробелом) и буквенные цепочки (n-граммы). Все три рода данных связаны друг с другом. Иллюстрацией станет табл. 2.1.

Таблица 2.1

	Относительная частота букв (в промилле)			
	Зона частых слов (0,000001 - 0,001)			Зона самых частых слов (>0,001)
	Начало слова	Внутри слова	Конец слова	
а	21	85	91	71
б	31	15	2	22
в	77	39	31	75
г	34	17	8	18
д	48	33	14	25
е	8	92	109	92
ж	7	10	1	20
з	28	15	5	15
и	34	81	98	98
й	...	4	85	-
к	59	31	12	40
л	19	46	18	26
м	49	23	63	22
н	64	77	16	76
о	67	117	91	109
п	148	13	1	20
р	54	65	12	16
с	122	58	11	45
т	26	75	50	68
у	24	22	31	13
ф	13	2	3	-
х	8	3	45	4
ц	6	8	1	-
ч	18	13	2	24
ш	5	7	...	-
щ	...	5	...	-
ъ	-	1	...	-
ы	-	17	43	19
ь	-	11	55	2
э	15	2	...	12
ю	2	4	21	-
я	5	9	80	13

¹ См. [Шайкевич 2008].

Анализ табл. 2.1 приводит к двум результатам.

1) Относительные частоты букв могут сильно варьировать в разных зонах частотного словаря. В зоне самых частых слов частоты букв *б, в, ж, и, к, ч, э* существенно выше, чем в зоне частых слов. Объяснение приходит сразу же — отдельные слова с максимальной частотой сильно влияют на общую частоту соответствующих букв. Таковы предлог *в*, союз *и*, формы глагола *быть*, местоимений *этот, какой, который, что* и т. п. Учет этого вывода заставит задуматься о способах расчета безусловной вероятности той или иной буквы.

2) Почти все буквы (за исключением буквы *у*) резко меняют свою частоту в разных позициях в слове: в начале слова видим громадное повышение частоты букв *н, с, в, к, б, э, ф*; понижена частота *о, и, т, а*; минимальна частота *е* и *й*, а буквы *ь, ы, ь* отсутствуют. В конце слова сосредоточены *й, я, м, ь, ы, х* и *ю*. Понятие позиции оказывалось ключевым в алгоритме Андреева. Условная вероятность какой-то буквы в данной позиции, деленная на безусловную вероятность, и была стартом при выявлении морфологических типов. Соответствующий коэффициент для конца слова был бы следующим: *й* = 5,7; *х* = 4,5; *я* = 4,2; *ю* = 3,5; *ь* = 3,4; *ы* = 2,4; *м* = 1,9.

На современных компьютерах можно получать статистику буквосочетаний заданной длины — *n*-грамм. Подобные данные использовались в ДСА еще до обращения к задаче открытия морфологии. Речь идет о собственно графематическом анализе — получении дистрибутивных классов букв (ср. 1.4.4), возможному открытию графем-диграфов. Последняя задача едва ли важна для русского языка, но она может стать ключевой для таких языков, как польский или английский. В английском языке ДСА, возможно, обнаружил бы группу диграфов, не ограниченных позицией в слове (*ch, ph, sh, th*), *wh*, сконцентрированный в начале слов, *gh*, упрятанный внутрь слова, и собственно *h*.

Таблицы буквосочетаний были бы здесь слишком громоздки. Ограничимся демонстрацией неслучайных значений *S* у буквосочетаний с *о* (*p* = 0,106), *н* (*p* = 0,031) и *р* (*p* = 0,056).

	S		S		S
оо	-1278	пп	-286	опр	+546
оп	-267	пр	+2908	орп	-149
ор	+356	ро	+828	пор	+340
по	+1294	рп	-504	про	+3487
		рр	-659	роп	+105
				рпо	-160

Ясно видны предпочитаемые буквосочетания и буквосочетания, встречающиеся ниже нормы.

Статистический анализ *n*-грамм может быть соединен с позицией. Взяв простейший случай диграмм, мы можем проверить степень концентрации данного буквосочетания в той или иной позиции. Так, у буквосочетания *ые* на конец слова приходится 99,6 % всех случаев. Вот перечень других диграмм, у которых в конечной позиции появляется более половины всех случаев (против ожидаемых 13 % в зоне частых слов): *-ии* 99,5 %, *-шь* 99 %, *-ый* 99 %, *-ых* 96 %, *-ях* 96 %, *-ию* 93 %, *-ой* 91 %, *-ся* 90 %, *-ют* 90 %, *-сь* 89 %, *-их* 87 %, *-ия* 84 %, *-ть* 83 %, *-ая* 81 %, *-ие* 77 %, *-ий* 77 %, *-ее* 74 %, *-ым* 73 %, *-ах* 69 %, *-ей* 69 %, *-ем* 67 %, *-ую* 67 %, *-ты* 64 %, *-ое* 55 %, *-им* 53 %, *-го* 52 %. Не столь высока концентрация диграмм в начале слова: *вз-* 95 %, *пр-* 80 %, *по-* 78 %, *вс-* 68 %, *за-* 57 %, *вы-* 55 %, *до-* 50 %. Подавляющее большинство этих диграмм войдут в состав парадигм, но есть здесь и информационный шум: конечное *-ты* и начальное *пр-*.

Что касается буквосочетаний внутри слова, то их интерпретация должна быть отложена до позднейших этапов открытия морфологии. Некоторые из них станут частью морфем или основ, например *-бяз-* (*S* = +1 370) в составе *обяз-*. Другие буквосочетания уже на этой стадии кажутся хорошими кандидатами в морфемы или комбинации морфем. Таковы *-ви-* (*S* = +300), *-йш-* (*S* = +460), *-уц-* (*S* = +690), *-юц-* (*S* = +2 100), *-ств-* (*S* = +4900). Наконец, в некоторых случаях внутрисловное буквосочетание может оказаться сигналом морфемных чередований: *-чн-* (*S* = +200), *-жн-* (*S* = +380), *-ьн-* (*S* = +1200).

Основное направление дальнейшей работы — анализ *n*-грамм как потенциальных основ и их расширений как потенциальных аффиксов. Попробуем ввести ограничения на длину основ и аффиксов. Пусть минимальная длина основы будет равной четырем буквам, а максимальная длина аффикса не превышает половины длины основы. Вводя такой порог, мы на этом этапе отказываемся от анализа многих коротких основ (вроде *дом-ами, мо-ими, взя-лись*) в надежде получить регулярную парадигматику, а уже затем переходить к обработке коротких основ. Удобный материал для проверки левых основ содержится в прямом алфавитном словаре. Соответствующий обратный словарь дает нам перечень правых основ. Дополнения до целого графического слова у левых основ станут потенциальными суффиксами, дополнения правых основ будут префиксами. Так, у левой основы *весн-* будут найдены следующие продолжения (суффиксы): *-а, -е, -ой, -у, -ы*, в своей совокупности образующие **протопарадигму**. В соответствии со с. 53 будем называть **широтой** протопарадигмы число аффиксов, входящих в ее состав (в данном случае —

5). Число основ, обладающих данным набором аффиксов, будем называть **мощностью** протопарадигмы. В нашем случае мощность парадигмы равна 30 (*анкар-, атмосфер-, блокад-, борьб-, генпрокуратур-, годовщин-, госдум-, дружб-, европ-, екатерин-, ирин-* и т. д.). В том случае, когда основа совпадает с целым графическим словом, но при этом обнаруживаются и какие-то аффиксы, отсутствие продолжения объявляется нулевым аффиксом (0). Например, парадигма *-0, -го, -е, -й, -м, -му, -сти, -сть* обладает мощностью 20 (*актуально-, вольно-, действительно-, исключительно- ...*). У парадигмы *-0, -а, -ом, -у* обнаружено 46 основ (*арафат-, белград-, борт-, вадим-, гайдар-, гейдар-, генсек-, голод-, гражданин-, дьявол- ...*). Чем шире парадигма, тем больше ее морфологическая определенность, тем больше у нее шансов совпасть с традиционным представлением о парадигме. Так, парадигма *-0, -й* при достаточной мощности (30 основ) включает очень разнородный материал (*безгранично-, богаче-, буре-, дружно-, иваново-, иронически- ...*). Заметим, что полученные парадигмы покрывают не только традиционно понимаемое словоизменение, но и словообразование. Особенно ясно это видно у префиксальных парадигм.

Префиксальные парадигмы отличаются минимальной шириной, но их мощность может быть весьма значительной. Вот важнейшие из них (с указанием мощности):

0, не	333	(велика, малое, которыми, уверенности ...);
0, по	222	(крова, важнее, иске, гибли, чести ...);
0, с	70	(высока, начала, можете, умели ...);
0, со	65	(стояла, братьев, хранение, жгли ...);
0, на	49	(всегда, завтра, звании, печатали ...);
0, за	48	(явка, хотела, будьте, купили ...);
0, у	48	(гроза, знала, видите ...);
0, гос	43	(сектора, чиновников, комитете, премий ...);
0, про	31	(демонстрировала, исходила, шествии ...);
0, рас	27	(продажа, скажите, положении ...).

Конечно, в полученных парадигмах присутствует информационный шум в виде пар правых основ, ошибочно объединенных в рамках одной парадигмы, ср.: *рема/зарема, града/награда, даре/ударе, лице/улице, грамма/программа, алии/реалии*, но доля таких ошибок невелика. Среди маломощных парадигм встречаются и полностью фиктивные, ср. парадигму *0-, 0-* с парами основ *остатка/достатка, вора/двора, оклад/доклад, ушами/душами, ревности/древности*. Однобуквенность префикса увеличивает вероятность ошибки. Уникален случай парадигмы *0-, с-*, где доля ошибок превышает 50 %, ср.: *танка/станка, тройка/стройка, тулу/стулу, цене/сцене*. Начальные консонантные сочетания со звуком *s-* столь обычны в индоевропейских языках, что приписывать им морфологический статус было бы рискованно.

Число суффиксальных парадигм очень велико. Десять из них имеют мощность более 199, у 52 парадигм мощность превышает 25 основ. Рекордсменом выступает парадигма *-0, -а* с 872 основами (*азот-, бахтин-, вальс-, глазунов-, дебют-, есенин-, живот-, зайцев-, имам-*). Особого внимания заслуживают две мощные парадигмы с единственным маркированным суффиксом: *-0, -ся* с 578 основами (*берут-, вернем-, встречать-, имеющими- ...*) и *-0, -сь* с 162 основами (*вернула-, готовили-, заняло- ...*). Оба суффикса встречаются в составе более широких парадигм, например, в парадигме *-0, -а, -ась, -и, -ись, -о, -ось, -ся* (*встречал-, высказал-, использовал-, изменил- ...*). Спаренность данных суффиксов, их графическое сходство позволяют считать их вариантами одной морфемы, присоединяемой к некоторым парадигмам.

Изредка полученные парадигмы точно соответствуют каким-то традиционным типам словоизменения. Такова парадигма *-0, -а, -ам, -ами, -ах, -е, -ов, -ом, -у, -ы* (*аспект-, вариант-, взгляд-, вопрос-, документ-, журнал-, институт- ...*) и почти точная *-0, -а, -ам, -ах, -е, -ой, -у, -ы* (*войн-, газет-, групп-, жертв- ...*).

Чаще полученные парадигмы представляют собой фрагменты традиционных парадигм, например, субстантивных: *-и, -ов* (*баск-, венк-, жулик-, значк- ...*); *-и, -ь* (*акварел-, важност-, ерес-, жидкост- ...*); *-а, -о* (*бешенств-, божеств-, гнезд-, древ- ...*); *-ей, -и, -ю, -я* (*абхазы-, бюрократи-, греци-, ирони- ...*); *-е, -ем, -и, -ю, -я* (*введени-, внимани-, давлени-, закрыти- ...*). Прообразами глагольных парадигм выступают *-0, -т* (*бегу, веря-, делаю-, ишу- ...*); *-л, -ла, -ли, -ть* (*висе-, воева-, вылете-, заговори- ...*); *-ет, -ли, -ть, -ют* (*влия-, добыва-, запреща-, исчеза- ...*). Наконец, примерами фрагментов адъективных парадигм служат *-ий, -ого* (*атлантическ-, белинск-, вольск-, жванецк- ...*) и *-ой, -ые, -ых* (*агитационн-, войсков-, заграничн- ...*).

В традиционном представлении адъективные парадигмы обладают наибольшей шириной, понятна поэтому малая вероятность найти в тексте полный набор соответствующих окончаний. Так, основа *иностранны-* в зоне частых слов имеет следующие продолжения: *-ая, -ого, -ое, -ой, -ому, -ую, -ые, -ый, -ым, -ыми, -ых*, для завершения парадигмы полного прилагательного не хватает лишь *-ом*. Основа *возглавляем-* имеет те же продолжения за исключением *-ыми*. Переход к зоне слов средней частоты привел бы к тому, что эти две основы оказались бы в пределах одной парадигмы.

Но где границы искомым парадигм? Ведь рядом с парадигмой основ типа *иностранн*- существуют парадигмы с дополнительными суффиксами *-а*, *-о*, *-ы* (*возможн*-, *адекватн*- и пр.). Расширение зоны поиска до слов средней частоты приводит к тому, что все больше основ проявляют сочетаемость с этими суффиксами (*избирательна*-, *временна*-, *глобальна*-, *денежно*-, *информационно*-). Более того, в парадигму включаются новые суффиксы: *бессмысленнее*-, *адекватнее*-, *естественней*-, *актуальней*-. Лишь суффикса *-ен* не хватает до полной традиционной парадигмы прилагательного (за вычетом суперлатива).

Существует еще одна мощная адъективная парадигма с исходом на *-ск*-, где на месте буквы *ы* в суффиксах мы находим букву *и* (*германск*-, *исламск*- и т. д.). Вместо суффикса *-о* обнаруживаем *-и* у основ *идеологическ*- и *демократическ*-. Переход к зоне средней частоты увеличивает круг подобных наречий (*европейски*-, *граждански*-), но мы не найдем здесь ни кратких прилагательных, ни компаратива. Попытка расширить парадигму оказалась бы тщетной, в русском таких форм нет.

Понижение порога частоты приводит к постоянному расширению парадигмы, наращиванию ее мощности, к слиянию разных парадигм. Если в зоне частых слов полная традиционная парадигма существительных среднего рода с исходом на *-ств*- ограничивалась 9 основами (*ведомств*-, *государств*-, *достоинств*-, *качеств*-, *министерств*-, *обществ*-, *правительств*-, *убийств*-, *чувств*-), то теперь их число выросло до 32 (*банкротств*-, *законодательств*-, *меньшинств*-, *существ*- и т. п.). Вместе с тем многие парадигмы не расширяются до теоретической полноты не из-за вероятностных ограничений, но вследствие каких-то семантических факторов. Действительно, именно семантика ответственна за то, что не выявлены формы множественного числа у таких слов, как *бегство*-, *большинство*-, *варварство*-, *взяточничество*-, *воровство*-, *господство*-, *дворянство*-, *детство*-, *духовенство*-, *единство*-, *животноводство*-, *изящество*-, *лидерство* и т. п.

Абстрактность существительного мешает его употреблению во множественном числе. Так, в зоне частых слов были обнаружены мощнейшая (хотя и узкая) парадигма с суффиксами *-и* и *-ь* (около 200 основ, большей частью с исходом на *-ост*) и близкая парадигма с суффиксами *-и*, *-ью* и *-я*. Соответствующая полная парадигма засвидетельствована только у основ: *возможност*-, *договоренност*-, *област*-, *обязанност*-, *особенност*-, *способност*-, *ценност*-. Понижение частотного порога приводит к переходу большинства основ из первой парадигмы во вторую, т. е. к достраиванию списка форм единственного числа. Но лишь 20 % из них пополняются формами множественного числа (*акварел*-, *банальност*- и т. п.).

В зоне частых слов выделилась парадигма с суффиксами *-ей*-, *-и*-, *-й*-, *-ю*-, *-я*-, *-ям*-, *-ями*-, *-ях*-, она включает 20 основ (*инстанции*-, *монополи*-, *ситуаци*-, *фотографи*- и т. д.). Понижение порога частоты резко увеличивает мощность парадигмы до 90 основ (*делегаци*-, *интерпретаци*-, *конституци*-, *презентаци*-, *цивилизаци*- и т. п.). Но некоторые основы с суффиксами *-ей*-, *-и*-, *-ю*-, *-я* отказываются строить формы множественного числа. В первую очередь это названия стран *абхаз*-, *араб*-, *англи*-, *албан*-, *венг*-, *грец*-, *ингушет*-, *индонез*-, *испан*- и т. д. К ним же примыкают и некоторые имена нарицательные: *авиация*-, *агитация*-, *документация*-, *интеллигенция*-, *инфляция*-, *ирония* и т. п.

Построению полных парадигм могут мешать исходно принятые ограничения на соотношение длины основы и длины аффикса. Например, в только что рассмотренной парадигме присоединению суффикса *-ями* препятствует краткость основ *лекци*-, *лини*-, *парти*-, *пенси*-, *преми*-, *сери*-, *сесс*-, *студи*-, *теори*-, *эмоци*-.

В связи с этим возникает еще один путь уточнения парадигм: принимая как промежуточный итог парадигмы, полученные в зоне частых слов, попробуем снять указанные ограничения на длину основы. В зоне частых слов с суффиксом *-ами* соединятся основы *автор*-, *актер*-, *выбор*-, *глаз*-, *завод*-, *игрок*-, *идеал*- и многие другие. У правой основы *-делил* к первоначальным префиксам *у*-, *вы*- присоединится *раз*-, а с понижением порога частоты также *0*-, *на*-, *по*-, *от*-. Соответственно у основы *-любил* к префиксам *0*-, *по*- прибавятся *раз*- и *воз*-. Уже в зоне частых слов левая основа *дава*- присоединит *-йте*-, *-лись* и *-лось* к прежним суффиксам *-й*-, *-л*-, *-ла*-, *-ли*-, *-ло*-, *-ть*-, *-я*-.

Оба пути уточнения парадигм (т. е. переход к зоне слов средней частоты и снятие ограничений на длину основ и аффиксов) сопровождаются появлением новых парадигм, связанных преимущественно со словообразованием. У вполне интегрированных в адъективную парадигму основ *вооруженн*-, *глобальн*-, *естественн*-, *избирательн*- появляются суффиксы *-ости* и *-ость*-, ведущие к субстантивным парадигмам. Субстантивный суффикс *-ик* обнаруживается у адъективных основ *духовн*- и *временн*-. К субстантивным основам присоединяются не только аффиксы адъективных парадигм: *институт*- + *-ские*-, *-ских*-, *-ской*-, *инвалид*- + *-ной*-, *-ном*-, *-ную*-, *-ные*-, *-ных*-, но и субстантивные аффиксы: *вопрос*-, *институт*- + *-ик*-, *газет*-, *голов*-, *групп*-, *икон*- + *-ка*-, *австри*- + *-йцы*-. В зоне частых слов основа *вспомина*- соединена с глагольными суффиксами *-ем*-, *-ет*-, *-ется*-, *-ешь*-, *-л*-, *-ла*-, *-ли*-, *-ть*-, *-ю*-, *-ют*-, *-ются*-, понижение частотного порога добавляет глагольный суффикс *-ете*-, адъективный *-виших* и субстантивный *-ние*-. Возникают особые субстантивные парадигмы (оппозиции мужского и женского) с суффиксами *-к*-, *-ка*-, *-ца*-, *-цы* (основы *великомучени*-, *волшебни*-, *воспитанни*-, *выпускни*-, *грешни*-, *заложи*-, *защитни*- и т. п.) и с суффиксами *-ь*-, *-я*-, *-ница*-, *-ницы* (основы *воспитател*-, *деятел*-, *жител*-, *зрител*- и т. п.). Постоянное «размножение» парадигм при пони-

жении порога частоты создает все новые и новые трудности, в общих чертах путь их преодоления представляется следующим.

В зоне частых слов формируются потенциальные автономные аффиксы из тех узких парадигм, где положительный аффикс противопоставлен нулевому. Для суффиксов такими внешними аффиксами кажутся *-сь* и *-ся*; среди префиксов наиболее яркий кандидат *не-*. Затем все графические слова частотного словаря переписываются заново уже без указанных внешних аффиксов («частиц»). Тем самым несколько сокращается широта и разнообразие парадигм, увеличивается их мощность.

Снова определяются парадигмы в обеих частотных зонах, но при этом особое внимание обращается на вид основ при аффиксах конкретной парадигмы. Если эти основы кончаются (или начинаются) на определенную букву (или комбинацию букв), то выдвигается гипотеза о компонентах этих основ, иначе говоря, о существовании основ разного порядка. Так, взяв только что приведенный пример с парадигмой *-ь, -я, -ница, -ницы*, находим все основы, соединяющиеся хотя бы с тремя из этих суффиксов. Таких основ оказалось 36, пять из них принадлежат к зоне частых слов: *житель-, исполнитель-, писатель-, представитель-, учитель-*, совпадая во всех четырех суффиксах. Сразу бросается в глаза асимметрия двух пар суффиксов. Частота суффикса *-ница* — 661, *-ницы* — 419, другие суффиксы из той же субстантивной парадигмы: *0* — 181, *-ам* — 22, *-ами* — 9, *-е* — 61, *-ей* — 149, *-у* — 90 (отсутствует суффикс *-ах* из полной парадигмы типа *больница*). Частота суффикса *-ь* — 27364, *-я* — 20308, другие суффиксы из той же субстантивной парадигмы: *-е* — 322, *-ей* — 28353, *-ем* — 7937, *-и* — 21835, *-ю* — 2943, *-ям* — 4466, *-ями* — 6720, *-ях* — 328 (т. е. полная субстантивная парадигма типа *автомобиль*). Последняя парадигма почти в 100 раз превышает первую по совокупной частоте.

От основ с исходом на *-ль-* создаются основы двух адъективных парадигм: 1) *зрительск-, издательск-, исполнительск-, любительск-, писательск-, предпринимательск-, председательск-, представительск-, преподавательск-, приятельск-, свидетельск-, учительск- и читательск-* с общей частотой 2 790; 2) *воспитательн-, деятельн-, зрительн-, исполнительн-, настоятельн-, основательн-, победительн-, представительн-, целительн-* с общей частотой 8 870.

От основ с исходом на *-льн-* образуются субстантивные основы *исполнительност-, основательност-, представительност-* и исключительно частая основа *деятельност-* (частота = 18 891)¹.

Непосредственно с основами на *-тель-* (основами первого порядка) соотносятся субстантивные основы с исходом на *-ств-* (основы второго порядка): *жительство-, издательств-, исполнительств-, любительств-, писательств-, покровительств-, предпринимательств-, представительств-, свидетельств-, сожительство-, учительств-* с совокупной частотой 9 188. Особое положение занимает основа *правительство-*, чья частота составляет 52 873. С ним соотносятся две адъективные основы третьего порядка: *правительственн-* (5 823) и *правительствующий* (25).

От некоторых основ с исходом на *-ств-* (*покровительств-, председательств-, представительств-* и *свидетельств-*) образованы по две глагольные основы третьего порядка с исходом на *-ова-* (с общей частотой 661) и с исходом на *-у-* (1 391), от последних созданы адъективные основы четвертого порядка с исходом на *-ищ-* (517).

Что касается самих основ на *-тел-*, то многие из них сами восходят к глагольным основам *исследова-, воспита-, жи-, изда-, исполни-, исследова-, люб-, надзира-, носи-, облада-, основа-, обита-, писа-, победи-, почита-, прави-, предпринима-, преподава-, представи-, руководи-, служи-, смотри-, созда-, учи-, храни-, цели-* и *чита-*. В остальных случаях явная первооснова в рамках ДСА не просматривается (*благодетел-, властител-, деятел-, зрител-, настоятел-, обывател-, покровител-, председател-, приятел-, свидетел-, сожител-*).

2.2. ОТ ПРОТОАФФИКСОВ К ПАРАДИГМАМ

2.2.1. ФОРМИРОВАНИЕ ПРОТОАФФИКСОВ И ПРОТОПАРАДИГМ

Существует множество путей статистического открытия морфологической структуры внутри графического слова. Тот путь, который был описан в предыдущем параграфе, учитывал частоту буквенных цепочек в корпусе текстов. Вернемся к нашему основному корпусу прозы и проверим еще один метод выявления парадигм. На этот раз зависимость от частот слов в тексте будет учитываться только в виде некоторого порога частоты, вводимого

¹ Заметим, что речь идет о формальных соотношениях основ: *деятель* — *деятельный* — *деятельность*. Реальное направление словопроизводства не может быть определено без привлечения семантических данных.

для ограничения списка графических слов, на котором будет работать программа формирования парадигм. Установление такого порога существенно сокращает вычислительную работу. Если в нашем корпусе встретилось 15 графических слов, начинающихся с буквенной цепочки *емельян* (*емельян, емельяна, емельяна-то, емельянихи, емельяниху, емельянов, емельянова, емельяновке, емельяновой, емельянову, емельяновы, емельяном, емельяну, емельянушка, емельяныч*), то лишь три слова преодолели установленный нами порог частоты (8): *емельян* — 61, *емельяна* — 22, *емельяну* — 9. Теперь *емельян* участвует в работе алгоритма на тех же правах, что и *иван*, частота которого в тексте в сто раз больше.

Снова устанавливаются пороги для длины основы (минимально 4 буквы) и для аффикса (максимально 4 буквы). Как только какая-то буквенная цепочка допускает хотя бы два продолжения в списке графических слов, начинается формирование разных основ. Начальная цепочка *емел* имеет только одно продолжение — *ь*, но цепочка *емель* допускает два продолжения — *к* и *я*, первое продолжение *емельк*, в свою очередь, допускает три продолжения: *а, и, у* (все они — конечные буквы графического слова). Цепочка *емельк* становится основой, а три продолжения объявляются протоаффиксами, их набор считается протопарадигмой. Ту же протопарадигму */а, -и, -у/* находим при основах *женк-, жэженк-, заметк-, зденк-, земляник-, золовк-, илюшечк-, кокетк-, конфетк-, копеечк-, косичк-, косушк-* и т. п. (всего — 71 основа). Несколько шире протопарадигма */а, -ами, -и, -ой, -у/* с основами *загадк-, занавеск-, игрушк-, нагайк-* (всего — 10 основ); еще шире протопарадигма */а, -ами, -е, -и, -ой, -у/* с основами *закуск-, копейк-, крышк-* (всего — 14 основ). Протопарадигма */а, -ам, -ами, -ах, -е, -и, -ой, -у/* с основами *выходк-, записк-, повозк-* почти совпадает с традиционно понимаемой парадигмой, отсутствует лишь форма род. п. мн. ч. (*-ок, -ек*). В рамках семантической лингвистики отождествление всех этих вариантов достигается моментально; в ДСА предстоит еще долгий путь.

Буквенная цепочка *емелья* допускает лишь одно продолжение *-н*, на этом, с одной стороны, завершается графическое слово и постулируется нулевой протосуффикс, а с другой — протосуффиксами становятся два других продолжения *а* и *у*. Возникшая протопарадигма */а, -а, -у/* относится к числу самых мощных — 177 основ.

Такой алгоритм породил более 10 тысяч протопарадигм, около 35 тысяч основ, 734 протоаффикса.

Результаты машинной обработки выдаются в следующих форматах:

1. Парадигмы с основами, например,

берег

1 0 а0 ам0 ами0 ах0 е0 и0 ись0 ите0 ла0 ли0 ов0 ом0 у0 ут0

правил

1 0 а0 ам0 ами0 ах0 е0 и0 о0 ом0 у0 ьяна0 ьнее0 ьно0 ьное0 ьной0 ьном0 ьною0 ьную0 ьны0
ьные0 ьный0 ьным0 ьных0

сапог

1 0 а0 ам0 ами0 ах0 е0 и0 ов0 ова0 ом0 у0

жених извозчик мальчик мужик начальник овраг помещик праздник старик успех
чиновник

11 0 а0 ам0 ами0 ах0 е0 и0 ов0 ом0 у0

бумаг собак

2 0 а0 ам0 ами0 ах0 е0 и0 ой0 ою0 у0

2. Частотный словарь протоаффиксов, например,

18201	0	37	ав0	14	аете0
26	а-то0	25	аем0	13	аешь0
6686	а0	226	ает0	58	ай0

3. Пары протоаффиксов, входящих в одни и те же протопарадигмы:

	Число общих парадигм	Число основ в этих парадигмах		Число общих парадигм	Число основ в этих парадигмах
0 а	1542	5152	0 ской	64	65
0 ам	273	338	0 съ	74	1519
0 ах	347	426	0 ся	162	2318
0 и	450	3723	0 те	158	496
0 ми	132	1456	0 у	798	1945
0 ов	494	859	0 ы	846	1235
0 ом	700	1176			

4. Значения S для пар протоаффиксов, входящих в одни и те же протопарадигмы:

S		S		S		S	
324	0 а	21	0 ли	37	0 ный	80	0 те
44	0 ам	280	0 м	25	0 ным	31	0 тся
62	0 ами	367	0 ми	27	0 ных	147	0 у
174	0 ась	66	0 му	77	0 о	304	0 х
49	0 ах	27	0 н	87	0 ов	32	0 ши
128	0 го	23	0 на	85	0 ом	31	0 ший
197	0 е	29	0 ная	120	0 ось	30	0 шка
49	0 есь	21	0 ного	60	0 сти	85	0 ы
318	0 и	21	0 ное	62	0 сть	24	0 ым
20	0 ии	33	0 ной	41	0 стью	71	0 ь
137	0 ись	20	0 ну	409	0 сь	44	0 ься
23	0 ия	21	0 ную	526	0 ся	73	0 ю
143	0 й	31	0 ные	82	0 т	35	0 я

Пары протоаффиксов могут быть ранжированы по значениям их S:

1695	ми х	388	е м	243	в м	198	ув ул
1069	ась ся	386	й ми	241	л ть	197	ь ься
1024	м х	367	0 ми	226	есь ься	196	вне вны
842	ась ись	364	го м	226	й ю	194	а ась
792	м ми	349	есь ся	220	го сть	194	ул уть
766	ись ся	328	го е	220	сти стью	193	ли ть
758	ась ось	328	сти сть	218	го сти	190	ула уть
730	й х	324	0 а	215	ул ула	188	ме ых
626	й м	320	е ми	212	сть стью	186	вне вну
589	ись ось	318	0 и	212	ся ься	186	ла ть
585	го му	304	0 х	210	и ись	185	ула ули
583	е х	290	а и	209	и ся	185	ме ый
537	го й	280	0 м	205	л ла	185	ь ью
526	0 ся	276	й му	202	те шь	184	и ось
517	ось ся	270	м му	201	ув ула	182	ув уть
512	е й	258	а у	200	ла ли		
409	0 сь	254	ась и	198	л ли		

2.2.2. ФОРМИРОВАНИЕ ГЛАГОЛЬНЫХ ПАРАДИГМ

Легко видеть, что уже на самом верху этого ранжированного списка пары протоаффиксов группируются в кластеры — обобщенные парадигмы. Четыре протоаффикса *-ась*, *-ись*, *-ось*, *-ся* самым естественным образом объединяются, ведь все они связаны друг с другом (среднее S = 755). Весь комплект этих четырех протоаффиксов представлен у 62 основ, одну из этих основ используем как ярлык для краткого обозначения соответствующей парадигмы, пусть это будет БОРОЛ-ся. Еще 159 основ сочетаются с тремя из данных четырех аффиксов. Наконец, около 50 основ соединились со всеми 4 данными аффиксами, но также с какими-то другими аффиксами. Все полученные основы кончаются на *-л* и, вероятно, вместе с другими основами образуют сложную глагольную систему. Заметим, что многие из этих основ всегда сопровождаются показателем *ся/сь* и образуют особый класс медиальных глаголов, ср.: *виднел*, *ворвал*, *годил*, *касал*, *мерещил*, *нравил*, *остал*, *осунул*, *откликнул*, *очутил*, *понадобил*, *появил*, *разгорал*, *смеял*, *столпил*, *толпил*, *улыбнул* и т. п.

Двигаясь вниз по ранжированному списку, мы обнаружим новые значимые связи: вышеприведенные аффиксы коррелируют с *0*, *-а* и *-и*. Создается парадигма ВОДИЛ, где 70 основ присоединяют все 7 аффиксов (и никаких других), а несколько сот основ сочетаются с 5—6 аффиксами из этого набора, либо присоединяют какой-то аффикс вне набора. Характерно сочетание рассматриваемых семи аффиксов с *-о* (полторы сотни основ типа *беспокоил-*, *вырвал-*, *потерял-*, *считал-* и т. п.). Последний протоаффикс не вошел в семичленный набор из-за многозначности русской морфемы *-о* (семантика среднего рода у существительных и форм прошедшего времени глагола, но также семантика наречия). Матрица значений S вновь присоединенных аффиксов и четырех аффиксов с *-ся/сь* (обозначена -СЯ) выглядит следующим образом:

	-ся	-о	-а	-и	-о
-ся	755	239	158	214	98
-о	239		324	318	77
-а	158	324		290	155
-и	214	318	290		99
-о	98	77	155	99	

Своеобразно поведение конечного *-ся*. Помимо того, что он входит в две предыдущие парадигмы (БОРОЛ-ся и ВОДИЛ), он появляется как конечный компонент в парадигме */-аяся, -еся, -ейся, -иися, -ихся, -уюся/ НАЧАВШ/КАСАЮЩ* (более 30 основ, среднее значение $S = 80$). Те же основы (без *-ся*) входят в парадигмы адъективного типа, но одновременно связаны с глагольными основами. Это заставляет признать *-ся* частицей или таким особым производным аффиксом, который не противостоит другим маркированным аффиксам¹. Подобное же поведение характерно и для префикса *не-*.

Две первые парадигмы содержали основы с конечным *-л*; это заставляет подозревать, что *-л* — какой-то суффикс, а наши парадигмы — лишь конечные компоненты в многослойной парадигме. Действительно, движение вниз по ранжированному списку S скоро обнаружит *-л* как протоаффикс в новой более широкой парадигме */-в, -ет, -ется, -л, -ла, -лась, -ли, -лись, -ло, -лось, -лся, -ть, -ться, -ют, -ются/*. Вот матрица попарных S внутри этой парадигмы²:

	-в	-ет	-ется	-л	-ла	-ли	-ло	-лся	-ть	-ться	-ют	-ются
-в		6	-	113	115	88	45	41	93	31	5	-
-ет	6		77	113	97	101	51	37	113	34	148	54
-ется	-	77		53	43	45	23	81	56	85	71	126
-л	113	113	53		205	198	97	79	241	72	97	37
-ла	115	97	43	205		200	116	71	186	63	86	32
-ли	88	101	45	198	200		109	69	193	62	99	35
-ло	45	51	23	97	116	109		46	92	36	46	18
-лся	41	37	81	79	71	69	46		125	80	128	35
-ть	93	113	56	241	186	193	92	80		82	98	41
-ться	31	34	85	72	63	62	36	128	82		32	76
-ют	5	148	71	97	86	99	46	35	98	32		57
-ются	-	54	126	37	32	35	18	74	41	76	57	

Бросается в глаза одно «аномальное место» в этой матрице — слабая связь *-в* с *-ет, -ется, -ют* и *-ются*, легко объясняемая с позиций семантической лингвистики: *-в* — это показатель прошедшего времени в деепричастиях, а *-ет* — флексия настоящего времени.

Нелегко подобрать ярлык для этой парадигмы. Нет ни одной основы, которая присоединяла бы весь набор аффиксов. Пусть представителем парадигмы будет НАЗЫВА-ть: основа *называет* присоединила 14 аффиксов (кроме *-в*) и еще 12 аффиксов, чья связь с рассматриваемой парадигмой обнаруживалась бы по мере снижения S^3 : *-ем* (12), *-емое* (12), *-емые* (15), *-емый* (15), *-емых* (10), *-ете* (40), *-еишь* (29), *-й* (18), *-йте* (36), *-ю* (29), *-юсь* (30), *-я* (49). Если рассмотреть все остальные обнаруженные протоаффиксы, то у 13 из них среднее S с аффиксами парадигмы НАЗЫВА-ть превысит 10: *-виная* (17), *-виши* (19), *-вишие* (22), *-виший* (29), *-виших* (17), *-емся* (14), *-йся* (20), *-тось* (14), *-ющие* (11), *-ющий* (16), *-ющим* (11), *-ющих* (13), *-ясь* (48).

Так постепенно определяются контуры того, что в семантической лингвистике называется глагольной парадигмой: выявились личные окончания, показатели инфинитива и деепричастий, наметились формы трех причастий. Исключительная широта складывающейся парадигмы (40 обнаруженных аффиксов) объясняет низкую мощность агрегируемых парадигм, причиной неполноты набора аффиксов у конкретной основы может быть ее малая частота. Самая мощная протопарадигма с 32 основами (*взира-, запуска-* и пр.) имеет только 4 общих аффикса (*-ет, -л, -ть, -я*); 14 общих аффиксов находим лишь у одной пары основ: *отвеча-* и *поступа-*. Заметим, что в рассматриваемой парадигме нет аффикса *-о*. Это указывает на связанный характер производящей основы (большинство основ заканчивается буквой *а*).

Спускаясь по ранжированному списку, мы, действительно, обнаружим парадигму, где основы уже не оканчиваются на *а*, напротив, у многих аффиксов это *а* обнаруживается. Данная парадигма ПРОЧИТ-ать включает */-ав,*

¹ С общелингвистической точки зрения *-сь* выступает как алломорф *-ся* после гласного во всех глагольных формах, кроме причастий.

² Под символом ЛСЯ объединены *-лась, -лись, -лось, -лся*.

³ В скобках указывается среднее S в связях аффикса с 12 аффиксами рассматриваемой парадигмы.

-аем, -ает, -аете, -ай, -айте, -ал, -ала, -али, -ался, -ан, -ана, -ано, -аны, -ать, -аю, -аюت, -ывай, -ывал, -ываю, -ывая/. Матрица внутренних связей свидетельствует о некоторой неоднородности парадигмы¹:

	АЕТ	АЛ	ать	АН	ЫВ
АЕТ	65	55	66	4	2
АЛ	55	141	120	36	18
ать	66	120		42	20
АН	4	36	42	73	28
ЫВ	2	18	20	28	78

С семантической точки зрения здесь объединился довольно разнородный материал как глаголы совершенного вида (*сказать, рассказать*), так и производные от них глаголы несовершенного вида, оформленные по-разному (*рассказывать, забывать, встречать*), отыменные глаголы несовершенного вида (*работать*).

С двумя последними парадигмами сходна парадигма СТРЕЛ-ять /-яем, -яет, -яете, -яй, -ял, -яла, -яли, -яло, -ялся, -ять, -яю, -яют, -яя/ со средним значением $S = 74$ у внутренних связей.

Общее конечное -н основы характеризуют парадигму ТОЛКН-уть /-ет, -ув, -ул, -ула, -ули, -уло, -ут, -уть/ со средним значением $S = 104$. Небольшой набор аффиксов обеспечивает достаточно мощную парадигму из шести основ: *блесн-, вздрогн-, вспыхн-, выглян-, дрогн-, толкн-*. Ясно, что должна появиться и парадигма типа НАМЕК-нуть /-нет, -нув, -нул, -нула, -нут, -нуть/ со средним значением $S = 53$.

Некоторым аналогом парадигмы НАЗЫВА-ть можно считать парадигму ПУСТИ-ть /-0, -в, -л, -ла, -ли, -ло, -м, -т, -те, -шь/, в которой подавляющее большинство основ кончается на букву *и*. Две самые мощные протопарадигмы содержат по 5 основ (*выпусти-, допусти-, отложи-, посади-, пропусти-* с 9 аффиксами и *кричи-, молчи-, посиди-, прогони-, услыши-* с 5 аффиксами).

В какой-то степени вариантом той же парадигмы можно считать парадигму ГОВОР-ить /-ив, -ил, -ила, -или, -ило, -ит, -ить, -ишь, -ят/ со средним значением $S = 85$.

Подобно тому, как в парадигме ПРОЧИТ-ать выявился компонент -ан, -ана, -ано, -аны, рядом с парадигмой ГОВОР-ить появляются парадигмы: ОПРЕДЕЛ-ЕН /-ен, -ена, -ено, -ены/ со средним значением $S = 85$ и ВЫСТАВ-ЛЕН /-лен, -лена, -лено, -лены/ со средним значением $S = 117$. Традиционная лингвистика моментально указала бы на закономерность: вариант с -л появляется после губных согласных. ДСА может лишь перечислить эти согласные (-б, -в, -м, -п), но не может обобщить результат, сказав, что это губные.

На парадигмы НАЗЫВА-ть и ПРОЧИТА-ть похожа парадигма СКРИП-еть /-ел, -ела, -ели/ со средним значением $S = 121$.

Особое положение занимает парадигма ЧУВСТВ-овал /-овал, -уем, -ует, -уете, -уй, -уют, -уя/ со средним значением $S = 64$.

Последняя парадигма, которую следует здесь упомянуть, включает полный вариант /-0, -ка, -те/ с 34 основами: *возьми, вспомни, выпей, выпьем, гляди, давай, изволь, налей, погляди, погоди, подай, подойди, подумай, поедем, пожалуй, позволь, пойдём, поживи, покажи, полно, попробуй, послушай, посмотри, постой, пошли, принеси, прочти, расскажи, скажи, слушай, смотри, спроси, ступай, сходи* и редуцированный вариант /-0, -ка/ с 14 основами: *велите, глянь, дайте, подите, подумайте, пойдёмте, пойдё, покажь, попробуйте, постоит, пусть, расскажите, садись, садитесь*.

Сходство рассмотренных до сих пор парадигм, проявляющееся в буквенном совпадении многих аффиксов, позволяет сформировать большой класс парадигм, который традиционная лингвистика назвала бы глагольным. Добраивание и уточнение парадигм будет производиться после снятия некоторых ранее введенных ограничений (порог частоты, пороги длины основы и аффикса), но черновое определение глагольного класса удалось получить в рамках графического слова. Вряд ли этот класс серьезно перестроится на минимальном интервале ДСА.

¹ Символом ЫВ обозначены четыре аффикса с этим буквосочетанием; соответственно используется символ АЛ; символ АН объединяет -ан, -ана, -ано, -аны и -ав; -ать показан отдельно; все остальные аффиксы объединены под символом АЕТ.

2.2.3. ФОРМИРОВАНИЕ АДЪЕКТИВНЫХ ПАРАДИГМ

Вернемся к началу ранжированного списка пар протоаффиксов. В пятерке первых пар две пары сыграли свою роль в формировании глагольных парадигм, оставшиеся три протоаффикса *-м*, *-ми* и *-х* будут встречаться во многих парадигмах, но никакой эксклюзивной¹ протопарадигмы они не создали. Восьмое и девятое место в ранжированном списке занимает *-й*, и тут же появляется первая адъективная трехчленная протопарадигма */-й, -м, -х/* с основами *артиллерийски-*, *верноподданны-*, *высочайши-*, *десяты-*, *еврейски-* ... всего 35 основ.

Опустившись до $S = 500$, обнаружим еще три протоаффикса *-го*, *-е*, *-му*, которым суждено войти во многие адъективные парадигмы. Появляется первая эксклюзивная протопарадигма с пятью протоаффиксами */-го, -е, -й, -м, -му/*: *английско-*, *барско-*, *ближайше-*, *будуще-*, *бывше-*, *военно-* ... всего 68 основ. Затем следует четырехчленная эксклюзивная протопарадигма */-е, -м, -ми, -х/*: *беспрестанны-*, *боковы-*, *висящи-*, *гнилы-*, *голубы-* ... всего 51 основа. Почти эксклюзивны пятичленные протопарадигмы */-го, -е, -й, -м, -му, -ю/*: *большо-*, *верхне-*, *весенне-* ... всего 45 основ; */-о, -го, -е, -й, -м, -му/*: *беспокойно-*, *блаженно-*, *божественно-*, *временно-*, *высочайше-*, *далеко-*, *душевно-*, *жарко-* ... всего 28 основ.

Дойдя до $S = 250$, увидим, что *-го* и *-ми* тесно связаны с протоаффиксом *-о*. Складывается первая агрегированная парадигма с восемью аффиксами */-о, -го, -е, -й, -м, -ми, -х/*, пусть ее ярлыком будет ГОЛОДНО. В составе этой парадигмы 396 основ с шестью аффиксами из данного набора, 666 основ с пятью аффиксами, 1 092 с четырьмя аффиксами. Было бы преждевременным считать ее адъективной, лучше назвать ее именной. Дело в том, что одна мощная парадигма */-о, -м, -ми, -х/* носит явно именной характер; из ее 270 основ лишь одна (*епанчины*) относится к адъективам, все остальные — субстантивы (*города, господа, греха, гусара, движения, двора* и т. п.).

Когда значения S достигают величины 120, множество протопарадигм легко агрегируется в широкую и мощную адъективную парадигму ТЯЖЕЛ-ый² */-ая (122), -ого (126), -ое (133), -ой (124), -ом (77), -ому (101), -ою (85), -ую (125), -ые (137), -ый (131), -ым (129), -ыми (111), -ых (121)/*. Тринадцать аффиксов создают такую широкую парадигму, что шансов обнаружить эксклюзивную парадигму практически нет. Основа ТЯЖЕЛ-, выбранная в качестве ярлыка, кроме этих 13 аффиксов присоединяют еще пять аффиксов *-а, -ее, -ей, -о, -ы*. Такой набор почти соответствовал бы полной парадигме традиционной лингвистики. Кроме *тяжел-* еще 89 основ сочетаются с этими 13 аффиксами: *благородн-, бледн-, богат-, болезненн-, быстр-, весел-, внезапн-, вольн-, главн-, громадн-, грустн-, грязн-, действительн-, деревянн-, длинн-, добродушн-, духовн-, единственн-, естественн-, законн-, замечательн-, значительн-, знаком-, известн-, интересн-, испуганн-, истинн-, казенн-, каменн-, котор-, красив-, красн-, кругл-, ласков-, любезн-, любим-, мертв-, мрачн-, народн-, наружн-, невинн-, недостойн-, неизвестн-, нектор-, необыкновенн-, неожиданн-, непонятн-, неприятн-, несчастн-, ничтожн-, нравственн-, обыкновенн-, обычн-, огромн-, особенн-, открыт-, печальн-, подобн-, полезен-, постоянн-, почтенн-, прекрасн-, приятн-, разбит-, раскрыт-, решительн-, светл-, свободн-, семейн-, серебрян-, серьезн-, сильн-, скверн-, скромн-, собственн-, спокоин-, старинн-, странн-, страстн-, страшн-, суров-, счастлив-, таинственн-, тверд-, толст-, холодн-, церковн-, честн-, чет-верт-*.

Любопытны внешние связи этой главной адъективной парадигмы с некоторыми внешними аффиксами: *-о* 90, *-ее* 61, *-ость* 55. Намечаются контуры более широкой парадигмы. Напомним, что в основе дистрибутивно-статистического выделения парадигм лежит идея регулярности (в данном случае — в пределах графического слова). Регулярные морфологические связи выходят за пределы традиционных представлений о частях речи. Если формы компаратива традиционно включаются в парадигму прилагательного, формы на *-о* попадают туда же лишь в функции предикатива, приглагольное или приадективное употребление объявляется особым словом — наречием. ДСАТ скорее включит все формы на *-о* в единую адъективную парадигму подобно тому, как деепричастия обычно включаются в единую глагольную парадигму.

Интересен вопрос о суперлативе. При заданных порогах на длину протоаффикса образовалась парадигма БЛАГОРОДНЕ-йший */-йшая, -йший, -йшим, -йших/*, не показавшая связей с адъективными парадигмами. В состав этой парадигмы входят 23 основы, сочетающиеся, по крайней мере, с двумя аффиксами из четырех: *благородне-, богате-, важне-, верне-, главне-, дальне-, добре-, искренне-, любезне-, огромне-, отличне-, перве-, полне-, почтенне-, превосходне-, прекрасне-, прелестне-, сильне-, совершенно-, старе-, счастливе-, ужасне-, чисте-*. Шестнадцать из них сочетаются с аффиксами *-ее* и *-о*. При допущении пятибуквенных аффиксов (и с переходом *-е* в аффикс) вся эта группа усилит свою мощь до 40 (*великодушн-, глуп-, изыщн-, подробн-* и т. п.) и целиком войдет в главную адъективную парадигму.

¹ Будем называть парадигму эксклюзивной, если она включает все те и только те аффиксы, которые явным образом перечислены в наборе данной парадигмы.

² В скобках указывается среднее S в связях аффикса с остальными аффиксами рассматриваемой парадигмы.

Следующая выделившаяся парадигма ХОРОШ-ий /-ая (74), -его (90), -ее (57), -ей (61), -ем (32), -ему (79), -ею (24), -ие (82), -ий (89), -им (91), -ими (85), -их (109), -ую (79)/ напоминает предыдущую хвостами аффиксов и заменой *ы* на *и*, *о* на *е*. Все 13 аффиксов представлены только у основы *горяч-*. Абсолютное большинство основ этой парадигмы — это шипящие согласные -ж-, -ш-, -ч-, -щ- (например, *будущ-, рабоч-, стоявш-, ближайш-, всеобщ-, похож-*). Напомним, что эти буквы не появляются перед -ы-, -ю-, -я-, следовательно, их групповая близость проявляется уже в графематическом анализе в таблице бинарных буквосочетаний. Впрочем, существуют 22 основы с исходом на -н с двумя заменами аффиксов -ая на -ая-, -ую на -ую: *верхн-, весенн-, вечерн-, внешн-, внутренн-, домашн-, древн-, замужн-, здешн-, излишн-, искренн-, крайн-, нынешн-, передн-, последн-, посторонн-, прежн-, соседн-, средн-, тамошн-, тогдашн-, утренн-*.

Еще одна сходная парадигма ВЫСОК-ий /-ая (97), -ие (70), -ий (81), -им (80), -ими (72), -их (91), -ого (105), -ое (85), -ой (91), -ом (63), -ому (80), -ою (98), -ую (104)/. Основы как правило в конце имеют -к (изредка -г). Графематические ограничения у -к и -г — те же, что и в предыдущей парадигме. Многие основы этой парадигмы (например, *близк-, крепк-, горьк-, сладк-, долг-, строг-*) к перечисленному набору добавляют -а-, -и-, -о-. Многими сотнями измеряется число основ с исходом на *ск*.

Разные исходы основ не изменяют исходов перечисленных адъективных аффиксов. С позиций семантической лингвистики такое поведение свидетельствует о существовании морфем. Рассмотрим в качестве примера случай парадигмы ХОРОШ-ий. Начнем с основ, заканчивающихся на -ш-. Формально выделяемый аффикс -ш-ая сочетается с основами *вы/пришед-, за/порос-, погиб-, привык-*, но в очень многих случаях (около 100) -ш-ая следует за буквой *в-*; значит, -вш-ая есть соединение суффикса -вш- и флексии -ая. Разобранный выше аффикс -йший также позволяет выделить флексию -ий, но где левая граница суффикса? -й- или -ей-? Что делать с основами *ближайш-, величайш-, высочайш-, глубочайш-, жесточайш-, кратчайш-, крепчайш-, мельчайш-, нижайш-, строжайш-, тончайш-*? Здесь открываются сложные правила морфемных чередований *к/ч, г/ж, лк/льч* и т. п. Парадигма /-ная, -ного, -ное, -ной, -ную, -ные, -ный, -ным, -ных/ допускает разбиение на -н- и -ая (ср.: *народ-, вздор-, комнат-, разум-, случай-, телес-*). Парадигма /-чная, -чно, -чное, -чной, -чную, -чные, -чный, -чным, -чных/ разбивается на -н-ая и предшествующую основу с исходом на -к/ч (ср. *восточный*), *г/ж* (ср. *бумажный*), *х/ш* (ср. *грешный*). Другое решение для случая -ьная, где *ь* отойдет к основе (ср.: *капитальная, правильная, похвальная*). Но что делать с /-нная, -нно, -нное, -нной, -нную, -нные, -нный, -нным, -нных/? Этим аффиксам всегда предшествует гласная буква, и значит, могут существовать -енная, -анная, -янная. Частично это отыменные прилагательные (*мужественная, бессмысленная, болезненная, письменная, деревянная*), частично — образования от глаголов парадигмы ГОВОР-ить (*благословенная, брошенная, возмущенная, вооруженная, воспитанная, выдуманная, вызванная, потерянная* и т. п.).

2.2.4. ФОРМИРОВАНИЕ СУБСТАНТИВНЫХ ПАРАДИГМ

Максимальные внутренние связи (среднее *S* = 253) показывает протопарадигма /-сти, -сть, -стью/, к которой явно примыкают /-стей, -стию, -стям, -стями, -стях/. Создается первая агрегированная парадигма с восемью аффиксами, к ней относится более двухсот основ, из которых лишь 8 имеют на конце *е* (*добле-, живуче-, пове-, поче-, преле-, свеже-, сове-, тяже-*), все остальные оканчиваются на *о*. Выберем в качестве ярлыка ОПАСНО-сть (с семью аффиксами); по 6 аффиксов у *благодарно-, должно-, обязанно-, способно-*, по 5 у *любезно-, местно-, неприятно-, подробно-, потребно-*. Почти все основы (98 %) входят в адъективные парадигмы.

Еще одну агрегированную парадигму образует тройка аффиксов /-ции, -цию, -ция/ со средним значением *S* = 142, к ней примыкает аффикс -тор (*S* = 91). Ярлыком парадигмы назовем АДМИНИСТРА-ция, четыре аффикса присоединяются также к основам *опера-, редак-, эску-*. По три аффикса находим у основ *агита-, депута-, дирек-, инквизи-, корреспонден-, планта-, провин-* (!), *протек-* (!), *репети-*.

Третья явная агрегированная парадигма («отчества») АЛЕКСЕЕ-вич имеет восемь аффиксов /-вич, -вича, -вичу, -вна, -вне, -вной, -вну, -вны/ со средним значением *S* = 142. Полностью восемь аффиксов видны еще у восьми основ: *александр-, андрее-, василье-, михайло-, николае-, платоно-, семено-, степано-*. Общее число основ не превышает 50. Внешние связи (среднее *S* = 52) с -ва и -ву ведут нас к парадигме («фамилии») НЕКРАС-ов /-ов, -ова, -ову, -овым/ со средним значением *S* = 69. Все 4 аффикса видны у основ *григор-, кирилл-, литвин-, меркул-, неведом-, неждан-, сверст-, степан-, стукол-, хвостик-, черемис-, целкал-*. Чаше встречаются основы с тремя первыми аффиксами: *базар-, васильк-, волын-, давид-, денис-, долгору-, ермол-, замет-, зосим-, колос-, кутуз-, лопух-, маркел-, молот-, молчан-, мороз-, найден-, налет-, новосел-, облом-, пален-, полоз-, раскольник-, розан-, рязан-, салтык-, семен-, сидор-, симон-, скворц-, стебельк-, устин-, федор-, христ-* (!), *чумак-*.

Многие сотни и даже тысячи основ принадлежат парадигме ОБРАЗ /-0, -а, -ам, -ами, -ах, -е, -ов, -ом, -у, -ы/, которая объединяет множество протопарадигм. Все десять аффиксов представлены у 33 основ: *взгляд, вопрос, генерал, диван, документ, журнал, закон, идеал, интерес, карман, класс, миллион, наряд, огород, ответ, офицер, переход, порыв, предмет, прием, пример, принцип, промысл, пункт, разговор, расчет, результат, роман, совет, стакан, столб, талант, француз*. По 9 аффиксов видим у основ: *берег, билет, вывод, выстрел, губернатор, гусар, забор, коридор, литератор, магазин, наряд, номер, обряд, остров, солдат, студент*. Внутренние связи аффиксов показаны в следующей матрице.

	-0	-а	-ам	-ами	-ах	-е	-ов	-ом	-у	-ы
-0										
-а	324		73	104	79	163	116	140	258	169
-ам	44	73		135	111	60	79	57	86	54
-ами	62	104	135		120	71	119	87	105	77
-ах	49	79	111	120		86	78	56	90	59
-е	197	163	60	71	86		53	66	174	82
-ов	87	116	79	119	78	53		112	83	85
-ом	85	140	57	87	56	66	112		100	66
-у	147	258	86	105	90	174	83	100		118
-ы	85	169	54	77	59	82	85	66	118	

Вариантом этой парадигмы можно считать набор аффиксов, где вместо -ы появляется -и: ЖЕНИХ /-0 (146), -а (172), -ам (68), -ами (98), -ах (87), -е (112), -и (139), -ов (88), -ом (83), -у (130)/. Эксклюзивной парадигмой здесь выступает набор основ: *жених, извозчик, мальчик, мужик, начальник, овраг, помещик, праздник, старик, успех, чиновник*. Как видим, многие перечисленные основы имеют исход на -ик. Естественно, что можно ожидать появление парадигмы ПРАЗДН-ик /-ик, -ика, -икам, -ике, -ики, -иков, -иком, -ику/ со средним значением $S = 81$. По восемь аффиксов имеют основы *начальн-, покойн-, священн-, чиновн-*; по 7 — *воспитанн-, преступн-, противн-, работн-, родственн-*; по 6 — *исправн-, любовн-, памяtn-, странн-*; по 5 — *истерик-* (явная ошибка!), *помещик-, посланн-, послушн-, пустыnn-, современн-, школьн-*. Некоторое сходство с парадигмой ПРАЗДН-ик демонстрирует парадигма ПОСЛОВ-иц /-иц, -ица, -ице, -ицей, -ицу, -ицы/ со средним $S = 105$. По шести аффиксов находим у основ *воспитанн-, гостин-, кормил-, красав-, любовн-, работн-, родственн-*; по пяти — *больн-, висел-, лестн-, маслен-, помещ-, сестр-*; по 4 — *мастер-, пугов-, стран-, странн-* (всего более 60 основ).

Хвосты аффиксов этой парадигмы ведут нас к чрезвычайно мощной (более тысячи основ) парадигме ПРИЧИН-а /-а, -е, -ой, -ою, -у, -ы/. Эксклюзивную парадигму находим в случае основ: *бездн-, белизн-, борьб-, быстрот-, вышиn-, глубин-, доброт-, дружб-, ивановн-, княжн-, корон-, николаевн-, нищет-, павловн-, простот-, пустот-, темнот-, теплот-, тишин-, толицин-, чистот-, шириn-*. Полный набор аффиксов наблюдаем также у 48 основ: *бесед-, бород-, величин-, грамот-, досад-, екатерин-, жалоб-, женицин-, жертв-, забот-, зашит-, звезд-, карет-, катерин-, комнат-, красот-, литератур-, минут-, могил-, молитв-, москв-, мужчин-, наград-, надежд-, натур-, невест-, оград-, перемен-, персон-, полнот-, половин-, польз-, посуд-, похвал-, правд-, природ-, просьб-, работ-, свадьб-, свобод-, сестр-, сигар-, служб-, сторон-, стран-, судьб-, угроз-, фигур-*.

«Мягкой» параллелью к парадигме ПРИЧИН-а можно считать парадигму КОНЮШ-ня /-не, -ней, -ни, -ню, -ня/: *барыш-, героин-, господ-, дерев-, колоколь-, пусты-, святы-, сплет-*.

Мощные парадигмы (ОБРАЗ, ЖЕНИХ, ПРИЧИН-а) — аналоги мужского и женского склонения традиционной русистики.

У аналогов слов среднего рода раньше всего обнаруживается парадигма ЗНАКОМ-ство /-а, -е, -о, -у/ со средним значением $S = 118$. По четыре аффикса находим у основ: *беспокой-, благород-, блажен-, государ-, дворян-, достоин-, наслед-, началь-, ничтоже-, обстоятель-, покровитель-, правитель-, производ-, простран-, самоубий-, совершен-, товарище-, устрой-, христиан-, человеке-*. Довольно слабы основания для выделения парадигмы ДЕРЕВ /-о, -ом, -у/. Все три аффикса сочетаются с основами: *болот-, войск-, гнезд-, зеркал-, колес-, кресл-, личик-, окошечк-, окошк-, письм-, правил-, ремесл-, стекл-* (не считая основ на -ств-).

Среднее значение $S = 114$ имеет парадигма ЗАКЛЮЧ-ение /-ение, -ении, -ений, -ению, -ения/. Все пять аффиксов наблюдаем у основ: *заключ-, наблюд-, направл-, наслажд-, объявл-, объясн-, определ-, оскорбл-, предлож-, предполож-, представл-, приглаш-, приготовл-, производ-, раздраж-, размышл-, разреш-, распорядж-, рассужд-, соображ-, стремл-*. Такое же среднее значение имеет парадигма ВОСПИТ-ание /-ание, -анию, -ания/, полностью представленная *образов-, оправд-, послуш-, предпис-, приказ-, содерж-, толков-*. Обобщением этих двух парадигм становится парадигма /-ние, -нии, -ний, -нию, -ния/ и еще дальше парадигма ДЕЙСТВ-ие /-ие, -ием, -ию, -ия, -иями, -иях/ со средним значением $S = 83$. Полный набор семи аффиксов видим также у 15 основ *воспомиан-, выражен-, движен-, намерен-, наслажден-, описан-, показан-, предан-, предположен-, рассужден-, сомнен-, сочи-*

нен-, страдан-, убежден-, увлечен-. Нет ничего удивительного в конечном *н* основ, поэтому ниже перечислим лишь основы с другим исходом: *занят-, извест-, несчаст-, понят-, услов-* (по 6 аффиксов); *предчувств-, прилич-, путешеств-, следств-* (по 5 аффиксов); *благополуч-, благород-, великодуш-, добродуш-, евангел-, милосерд-, насил-, невер-, открыт-, отсутств-, препятств-, приветств-, присутств-, проклят-, простодуш-, противореч-, развит-, разнообраз-, соглас-, сочувств-, усерд-* (по 4 аффикса).

Уже при значении $S = 128$ обнаружена пара *-ье* — *-ья* (*бездел-, богомол-, варень-, вязан-, гулян-, жалован-, забыт-, зареч-, кушан-, мужич-, нездоров-, ненаст-, печен-, повер-, подпол-, помест-, предмест-, раздум-*)¹. Теперь выясняется, что эта пара ассоциируется с парадигмой *ДЕЙСТВ-ие* (среднее $S = 100$), самые характерные основы *воспоминан-, движен-, намерен-, наслажден-, рассужден-, сомнен-, страдан-* получают теперь вариант с *-ье*. Возрастает и число доказанных основ: *бесчест-, весел-, воскресен-, лечен-, прощан-, сиден-, соглас-*.

Среднее значение $S = 128$ показала одна узкая парадигма, соответствующая словам среднего рода, *ЧУДОВ-ище /-ища, -ище/*: *обитал-, пепел-, пожар-, пристан-, прозв-*.

Около $S = 100$ в ранжированном списке появляется четырехчленная парадигма *СОЧИНИ-тель /-тели, -тель, -телю-, -теля/*, полностью представленная у основ *заседа-, покупа-, посети-, представи-, следова-, смотри-, управи-*.

Обобщением этой парадигмы можно считать *ПРОФИЛ-ь /-ем, -ь, -я/* с эксклюзивной парадигмой: *вензел-, мишел-, обладател-, поручител-, ридикул-*. Все три аффикса сочетаются также с основами *богатыр-, гвозд-, карас-, крендель-, пастыр-, спектакл-, тюлен-, щегол-*.

Опускаясь дальше по ранжированному списку, впервые сталкиваемся с чередованием букв внутри формирующихся парадигм. Примером может служить парадигма *БУТЫЛ-ка/ЖЕЛУД-ок/-ка, -ке, -ки, -ку, -ок/* со средним $S = 102$. Все пять аффиксов встретились у 49 основ: *арестант-, бесед-, бутыл-, волос-, вывод-, выдум-, выход-, голов-, город-, гувернант-, дворян-, десят-, желуд-, закус-, запис-, казач-, картин-, короб-, крестьян-, лошад-, малют-, мецан-, недостаток-, останов-, палат-, перелес-, переул-, перчат-, повоз-, подбород-, поезд-, покуп-, попыт-, поступ-, прогул-, продел-, распис-, решет-, сирот-, скрип-, служан-, сосед-, справ-, старич-, стрел-, тарел-, тетрадь-, уступ-, цыган-*. По 4 аффикса у основ: *англичан-, бабен-, беспоряд-, вывес-, задат-, замет-, занавес-, корот- (!), мужич-, налив-, напиток-, недоим-, остат-, отрыв-, перед-, перепел-, песен-, печен-, подар-, поедин-, поправ-, посел-, постав-, посыл-, потол-, пригор-, припад-, промежут-, проступ-, рассуд-, склад-, товар-, торгов-, участ-, холод-, шарман-;* по 3 аффикса — *берез-, богомол-, булав-, ботин-, голуб-, достат-, дурач-, жаворон-, заработ-, избыт-, институт-, кредит-, образ-, околот-, отговор-, перекрест-, петров-, пирож-, потом-, почин-, приживал-, просел-, сверт-, червяч-*.

Данная парадигма связана одновременно с парадигмами *ЖЕНИХ* и *ПРИЧИН-а*, что проявляется в ее внешних связях с аффиксами *-кам, -ками, -ках, -ков, -кой, -ком* (в среднем $S = 56$).

Близок случай парадигмы *НАСМЕШ-ка /-ек, -ка, -ке, -ки, -ку/* с основами *бабуш-, бумаж-, верхуш-, девоч-, девуш-, дорож-, карточ-, лавоч-, мальчиш-, насмеш-, отлуч-, палоч-, подуш-, привыч-, рубаш-, собач-, старуш-, строч-, фураж-*. По 4 аффикса у основ *денеж-, игруш-, казач-, клетуш-, кусоч-, местеч-, платоч-;* по 3 — у *годоch-, листоч-, человек-*. Аффикс *-ек* связан ($S = 33$)² с четырехчленной парадигмой *ВАСЕН-ька /-ька, -ьке, -ьки, -ьку/* (среднее $S = 94$), где по 4 аффикса имеют основы *варен-, дяден-, полин-, наден-, постел-;* по 3 — *барын-, деревен-*.³

На парадигмы *ЖЕЛУДОК* и *НАСМЕШ-ка* похожа парадигма *СКВОР-ец /-ец, -ца, -цы/* (среднее $S = 96$): *американ-, безум-, выход-, глупов-, голишин-, гостин-, жереб-, живопис-, знаком-, иностран-, испан-, итальян-, колод-, красав-, любим-, мерзав-, младен-, молод-, незнаком-, образ-, ополчен-, ординар-, очевид-, питом-, полковод-, послан-, преображен-, продав-, проходим-, румян-, сослужив-, счастлив-, ташкент-, торгов-, червон-, чечен-*.

Подобно парадигме *ВАСЕН-ька* засвидетельствована парадигма */-ьца, -ьцев, -ьцы/* (среднее $S = 86$) с тремя основами *владел-, земледел-, постоял-*.

С семантической точки зрения 25 основ, относящихся к парадигме */-ин, -ина, -ину/* (среднее $S = 103$), должны распасться, по крайней мере, на три кластера: 1) *велич-, развал-, стари-*; 2) *бабушкин-* и многие фамилии: *истом-, красотк-, ракич-, рогож-, рябин-, солом-* и др.; 3) *англичан-, господ-, гражд-ан-, дворян-, крестьян-, мецан-, татар-*. Формальное представление об этом дает матрица связей кластеров с некоторыми протоаффиксами.

¹ Но здесь же встретились основы *игумен-, настас-, наталь-, попад-, устин-*; *девич-, охотнич-, собач-, человек-*, которые не обнаружат общих аффиксов с парадигмой *ДЕЙСТВ-ие*.

² Примером соответствующих основ служат *васил-, ступен-, козыр-, кошел-, флигел-*.

³ Преобладание *н* в исходе этих основ заставит думать о суффиксе *-ньк-*, эта гипотеза подтвердится после снятия ограничения на длину основ и аффиксов (см. п. 3.2.5), ср. субстантивные основы *агашеньк-а, алешенька, анниньк-а, афоньк-а, бабеньк-а, батеньк-а, васеньк-а* и др.; адъективные основы *аленьк-ий, бедненьк-ий, беленьк-ий, веселеньк-ий* и т. п.

	Кластер 1	Кластер 2	Кластер 3
-иной	66	5	—
-ины	170	65	—
-ами	—	—	68
-е	—	15	70

С формальной точки зрения особой парадигмой (хотя и с низким значением S) следует признать объединение двух парадигм с 12 аффиксами /-ская, -ские, -ский, -ским, -ских, -ское, -ской, -скую, -ства, -стве, -ство, -ству/. Среднее значение в первой парадигме $S = 120$, во второй — $S = 118$, связи двух парадигм имеют среднее $S = 20$. Число соответствующих основ невелико¹, но все они крайне любопытны семантически — *христиан-*, *человече-* (12), *дворян-* (11), *господ-*, *граждан-*, *старче-* (10), *щеголь-* (9), *мастер-* (8), *варвар-* (6), *королев-*, *плутов-* (5).

Подведем итог пункту 2.2. Список пар протоаффиксов, ранжированный по величине S, т. е. по степени статистической неслучайности, стал основой для кластеризации протопарадигм. На множестве десяти тысяч протопарадигм сформировались 43 агрегированные парадигмы. Ниже дается их список, где слева указан ярлык, а справа — аффиксы парадигмы:

Глагольные парадигмы

1. НАЗЫВА-ть /в, вшая, вши, вшие, вший, вших, ем, емое, емся, емые, емый, емых, емся, ет, ете, ется, ешь, й, йся, йте, л, ла, лась, ли, лись, ло, лось, лся, тесь, ть, ться, ю, юсь, ют, ются ющие, ющий, ющим, ющих, я, ясь/
2. ВОДИЛ /0, а, и, о/
3. ПРОЧИТ-ать /ав, аем, ает, аете, ай, айте, ал, ала, али, ался, ан, ана, ано, аны, ать, аю, ают, ывай, ывал, ываю, ывая/
4. СТРЕЛ-ять /яем, яет, яете, яй, ял, яла, яли, яло, ялся, ять, яю, яют, яя/
5. ПУСТИ-ть /0, в, л, ла, ли, ло, м, т, те, шь/
6. ГОВОР-ить /ив, ил, ила, или, ило, ит, ить, ишь, ят/
7. ОПРЕДЕЛ-ен /ен, ена, ено, ены/
8. ВЫСТАВ-лен /лен, лена, лено, лены/
9. ТОЛКН-уть /ет, ув, ул, ула, ули, уло, ут, уть/
10. НАМЕК-нуть /нет, нув, нул, нула, нут, нуть/
11. СКРИП-еть /ел, ела, ели/
12. ЧУВСТВ-овал /овал, уем, ует, уете, уй, уют, уя/
13. ГЛЯДИ /-ка, 0, те/
14. БОРОЛ-ся /ась, ись, ось, ся/
15. НАЧАВШИЙся/КАСАЮЩ-ийся /аяся, еesia, ейся, иesia, ийся, ихся, уюся/

Адъективные парадигмы

16. ГОЛОДНО /0, го, е, й, м, ми, х/
17. ТЯЖЕЛ-ый /а, ая, ее, ей, о, ого, ое, ой, ом, ому, ою, ую, ы, ые, ый, ым, ыми, ых/
18. ХОРОШ-ий /ая, его, ее, ей, ем, ему, ею, ие, ий, им, ими, их, ую /
19. БЛАГОРОДНЕ-йший /йшая, йший, йшим, йших/
20. ВЫСОК-ий /ая, ие, ий, им, ими, их, ого, ое, ой, ом, ому, ою, ую/

Субстантивные парадигмы

21. ОБРАЗ /0, а, ам, ами, ах, е, ов, ом, у ы/
22. ЖЕНИХ: /0, а, ам, ами, ах, е, и, ов, ом, у/
23. ПРАЗДН-ик /ик, ика, икам, ике, ики, иков, иком, ику/
24. СКВОР-ец /ец, ца, цы/
25. ПРОФИЛ-ь /ем, ь, я/
26. СОЧИНИ-тель /тели, тель, телю, теля/
27. ПРИЧИН-а /а, е, ой, ою, у, ы/
28. КОНЮШ-ня /не, ней, ни, ню, ня/
29. ПОСЛОВ-иц /иц, ица, ице, ицей, ицу, ицы/

¹ В скобках указано число обнаруженных аффиксов.

Часть 3

**Частотный словарь языка русской прозы
1850—1870-х гг.**

О ЧАСТОТНОМ СЛОВАРЕ

Частотный словарь построен на основе корпуса, включающего тексты 52 авторов общим объемом около 15 миллионов словоупотреблений.

В нижеследующем списке для каждого автора указан объем текстов (в тысячах словоупотреблений). Тексты, вошедшие в корпус, брались целиком. Единственное исключение — произведения А. К. Шеллера-Михайлова, взятые как совокупность выборок (каждая по три страницы книжного издания).

М. В. Авдеев	76	А. О. Новодворский	37
Д. В. Аверкиев	21	А. Н. Островский	523
С. Т. Аксаков	330	И. И. Панаев	230
П. Д. Боборыкин	114	А. Ф. Писемский	1299
В. М. Гаршин	30	Н. Г. Помяловский	126
А. И. Герцен	474	А. А. Потехин	270
И. А. Гончаров	634	Ф. М. Решетников	153
И. Ф. Горбунов	138	Е. А. Салиас	188
Н. И. Греч	134	И. А. Салов	79
Д. В. Григорович	151	М. Е. Салтыков-Щедрин	838
А. А. Григорьев	109	В. А. Слепцов	136
Г. П. Данилевский	147	В. А. Соллогуб	24
Ф. М. Достоевский	1649	Вс. С. Соловьев	128
Ю. В. Жадовская	104	К. М. Станюкович	292
В. Г. Короленко	14	А. В. Сухово-Кобылин	51
В. В. Крестовский	625	А. К. Толстой	91
П. В. Кукольник	18	Л. Н. Толстой	1184
А. И. Левитов	30	И. С. Тургенев	697
Н. А. Лейкин	86	Г. И. Успенский	265
К. Н. Леонтьев	326	Н. В. Успенский	103
Н. С. Лесков	1423	А. А. Фет	43
С. В. Максимов	132	И. Е. Чернышев	8
П. И. Мельников-Печерский	677	Н. Г. Чернышевский	139
Д. Л. Мордовцев	285	А. П. Чехов	54
Н. И. Наумов	39	А. К. Шеллер-Михайлов	115
Вас. И. Немирович-Данченко	65		
И. С. Никитин	26		
		==	14896

В исследуемом корпусе преобладает художественная проза (романы, повести, рассказы, очерки — более 87% общего объема корпуса). Примерно 4% корпуса составляют драматические произведения, написанные прозой. Включены воспоминания Греча, Григорьева, Панаева, «Былое и думы» Герцена, «Фрегат “Паллада”» Гончарова, «Записки об уженье рыбы» и «Записки ружейного охотника Оренбургской губернии» Аксакова. Авторская, жанровая и хронологическая дифференциация лексического материала будет показана во втором томе.

Словник. В словаре представлены слова с частотой 3 и более. Имена собственные (личные имена, фамилии реальных лиц, общеизвестные топонимы, названия общеизвестных произведений, названия газет и журналов) включаются в словарь, начиная с частоты 10. Тот же порог частоты принят для чисел в цифровой записи. Отчества и фамилии персонажей в словарь не включены.

Слова, записанные латиницей, поданы отдельно. В основном корпусе словаря иноязычные слова, записанные кириллицей, сопровождаются следующими указаниями языка: А — английский, Е — (древне)еврейский, И — итальянский, Л — латинский, Н — немецкий, П — польский, У — украинский, Ф — французский, Ц — церковнославянский.

Леммы. В настоящем словаре помимо общепринятых лемм довольно широко представлены сочетания слов как особые словарные статьи, ср.

493	в качестве	684	в течение
838	в особенности	180	в упор
23	в пику	487	за границей
643	в продолжение	664	за границу
244	в сопровождении	178	за исключением
695	в сущности	109	за неимением

Частота таких словосочетаний не включена в частоту слов-компонентов, но при словах-компонентах дается отсылка к словарной статье словосочетания, примером служит словарная статья

550 качество
 22 нравственные к.
 [тж. в качестве] .

Читатель, интересующийся общей частотой слова *качество*, произведет арифметическую операцию $550 + 493 = 1043$.

В некоторых случаях словарную статью глагола возглавляет не инфинитив, а какая-то другая форма. Наряду со статьями

122 сдаваться,

в словаре есть и статья

172 сдается (-лось) .

Ср. также

1872 стоит (-ило)
 624 не стоит
 11 не стоит обращать
 9 стоило взглянуть;
 1243 стоять
 42 дорого стоять
 14 не стоит благодарности
 4 не стоит свеч.

Отдельными леммами считаются формы сравнительной и превосходной степени.

Разведение омонимов сопровождается минимальной семантизацией (обычно только при одном из них), ср.

54 балка
 11 балка (овраг) ;
 10 газовый
 3 газовое платье
 36 газовый (хим.)
 13 газовый рожок;
 116 град
 15 пот градом
 120 град (город) ;
 310 жать
 52 жать (жнет) ;
 752 коса
 15 густая коса
 9 распущенная коса
 9 русая коса
 13 седая коса
 25 черная коса
 86 коса (с-х)
 7 нашла коса на камень
 30 коса (земли)
 10 песчаная коса;
 62 лук
 91 лук (овощ) .

Словарь включает более 51 тысячи лемм.

Структура словарной статьи. Статья открывается цифрами (абсолютной) частоты, за ней следует лемма. Если в статье есть еще какая-то информация, то все нижеследующие строки даются с отступом (соответствующие частоты уже включены в частоту при лемме). В таких строках могут даваться некоторые грамматические формы, ср.

187729 быть
 11630 быть (инф.)
 22653 будет
 731 будучи
 2358 будь
 81 быв
 96 бывши

132939 был
16985 есть
247 суть¹

Частота формы второго лица глагола всегда включает и форму множественного числа. Это правило действует и тогда, когда форма второго лица фигурирует как особая лемма, ср.

1694 позволить
611 не позволить
2977 позволь;
132 здравствовать
50 да здравствует
1586 здравствуй.

Внутри словарной статьи могут приводиться какие-то варианты леммы, ср.

394 Ардальон
47 Ардалион

Если такие варианты носят простонародный характер, они даются курсивом:

152 артиллерия
6 антиллерия;
55 вошь
5 вша.

Отличительная черта данного словаря — широкое включение бинарных словосочетаний в состав словарных статей.

7007 слушать
31 жадно слушать
6 любопытно слушать
43 любил слушать
1314 слушай
42 слухать

Чаще всего бинарное словосочетание из прилагательного, согласуемого с последующим существительным, ср.

1478 зеленый
13 зеленая скамейка
10 зеленая трава
15 зеленое платье
19 зеленое поле
36 зеленое сукно
19 зеленый цвет

Внутри таких бинарных словосочетаний пунктуация не дается, ср.

7675 ах
481 ах боже
66 ах господи
410 ах да
498 ах как
221 ах нет
863 ах ты
54 ах черт

Символ // указывает на возможность двойного порядка слов внутри бинарного словосочетания:

74 битком // набитый
97 сердце // бьется
975 Бога // ради

¹ За формой *будет* скрываются также *буду, будешь, будем, будете, будут*; за формой *был* — соответственно — *была, было, были*.

Таблица 3.1

А

140945	а	59	Авенир	426	адвокат
4208	а то	541	авось	18	аблакат
1992	а (междом.)	14	авось-либо	3	абвокат
1956	а (вопрос)	25	Авраам	3	адвокатишка
22	а (буква)	10	Авраамий	23	адвокатский
10	а (звук)	4	авраамов	9	адвокатство
453	А.	9	авральный	17	адвокатура
21	а ля	14	Аврора	312	Аделаида
39	абажур	5	австралийский	3	аделаидин
28	Абакум	26	Австралия	22	Адель
70	аббат	48	австриец	15	адепт
5	аббатик	270	австрийский	200	административный
11	аббатство	112	Австрия	119	администратор
7	абвахта	15	австрияк	3	администраторский
14	Абдул	7	австриак	127	администрация
58	Абдулка	8	автобиография	5	администрировать
3	абие (Ц)	13	автомат	344	адмирал
12	або	9	автоматически	27	адмиралтейский
4	аболиционист	4	автомедон	15	Адмиралтейская площадь
10	абонемент	694	автор	26	адмиралтейство
11	абонированный	295	авторитет	29	адмиральский
11	абонироваться	25	авторитетно	6	а. час
6	абордаж	24	авторитетный	167	адмиральша
10	абориген	33	авторский	13	адов
3	абракадабра	5	авторство	30	Адольф
200	Абрам	179	ага	16	Адонирам
13	Абрамка	5	ага (сущ.)	4	адоратер
48	абрек	78	Агап	428	адрес
6	абрикос	31	Агапия	7	адрес-календарь
4	абрикосовый	6	агаряне	24	адресный
9	абрис	4	агат	71	адресовать
13	абсент	62	Агата	41	адресоваться
3	абсинт	4	агатовый	34	Адриан
3	абсентеизм	24	Агатон	67	Адрианополь
3	абсолютизм	97	Агафон	13	адрианополюский
12	абсолютно	354	Агафья	4	адриатический
65	абсолютный	79	Агаша	17	адски
5	абстрактность	11	Агашенька	162	адский
9	абстрактный	52	Агашка	12	а. мука
7	абсурд	802	Аггей	5	а. камень
3	абцуг	158	агент	5	адъюнкт
7	абшид	12	агентство	683	адъютант
10	абы	7	агентша	8	полковой а.
4	аван-ложа	37	агитатор	6	адъютантик
32	авангард	26	агитация	10	адъютантский
3	авангардный	9	агитировать	14	адье (ф)
12	аванпост	490	Аглая	7	адью
4	аванс	61	агнец	6	адюльтер
79	авансцена	110	Агния	7	аер
28	авантаж	3	агон	35	аж
3	авантажнее	47	агония	8	ажю
4	авантажно	9	предсмертная а.	5	ажютаж
7	авантажный	4	аграрный	23	ажитация
5	авантюра	701	Аграфена	3	ажитировать
4	авантюрист	4	аграфенин	62	ажно
18	авантюристка	36	Агриппина	12	ажурный
84	Аввакум	25	агроном	251	аз (Ц)
10	авгур	13	агрономический	24	аз (буква)
500	август	19	агрономия	190	азарт
33	Августа	4	агу	21	азартно
13	августейший	3	агунюшки	3	азартность
19	Августин	302	ад	43	азартный
11	Августина	5	исчадие ада	98	азбука
4	августинец	124	Адам	4	азбучка
3	августинский	18	адамант	9	азбучный
14	августовский	3	адамантовый	21	азиат
16	Авдей	24	адамов	6	азиат
15	Авдотьяшка	15	а. голова	85	азиатский
848	Авдотья			7	азиатский
16	Авель			3	азиатцы

50	Азия	382	Акулина	10	Аллан
12	Азов	36	Акулька	17	Аллах
14	азовский	6	акулькин	5	аллегорически
5	азот	18	Акундин	11	аллегорический
3	азотный	13	акурат	65	аллегория
17	азы	3	акустический	9	аллейка
7	азям	30	акушер	523	аллея
3	аир	127	акушерка	28	липовая а.
17	аист	6	акушерский	39	темная а.
505	ай	6	акушерство	5	тенистая а.
32	ай батюшки	98	акцент	4	уединенная а.
200	ай да	14	немецкий а.	20	аллилуйя
81	ай-ай	3	парижский а.	4	сугубая а.
10	ай-вай	8	польский а.	14	аллюр
30	ай-люли	3	акциденция	44	алмаз
7	айва	10	акциз	41	алмазный
51	айда	7	акцизник	5	алость
32	академик	13	акцизница	14	алоз
44	академический	3	акцизничиха	210	алтарь
282	академия	3	акцизно-откупной	68	алтын
16	а. наук	34	акцизный	8	алтынник
22	акафист	44	акционер	6	алфавит
96	акация	46	акционерный	3	алфавитный
4	аквамариновый	152	акция	6	алхимик
19	акварель	5	алатырь	3	алхимический
15	акварельный	5	албанский	3	алхимия
13	акварий	3	алгвазил	3	алчно
4	акведук	22	алгебра	35	алчность
139	аки	7	алгебраический	37	алчный
419	Аким	33	але (ф)	223	алый
4	акклиматизация	10	алебарда	5	альбатрос
4	акклиматизировать	19	алебастр	126	Альберт
4	акклиматизироваться	10	алебастровый	10	Альбион
44	аккомпанемент	15	Алей	28	альбо (П)
4	компамент	3185	Александр	148	альбом
3	аккомпаниатор	19	Лександр	5	альбомчик
59	аккомпанировать	11	Ликсандра	4	алькад
119	аккорд	51	Александр I	23	альков
31	аккурат	28	Александр Македонский	11	альмавива
16	аккуратнее	413	Александра	48	альманах
9	аккуратнейший	20	александрийский	10	альпийский
270	аккуратно	18	Александрийский театр	14	Альпы
100	аккуратность	14	Александринский театр	22	альт
106	аккуратный	13	Александрия	3	альтернатива
4	акр	41	александровский	9	альфа
15	акриды	9	а. лента	14	Альфонтс
27	акробат	40	Алексаша	30	Альфонсина
4	акробатический	25	Алексашка	42	Альфонсинка
8	акростих	4233	Алексей	14	Альфонсо
236	Аксаков	6	Лексей	40	Альфред
3	аксамитный	9	Олексей	3	альянс
44	аксельбант	53	Алексеюшка	5	алюминиевый
5	эксельбанты	68	Алексис	7	алюминий
53	Аксен	177	Алена	8	аляповато
54	Аксентий	16	аленин	23	аляповатый
17	аксесуар	59	Аленка	56	амазонка
91	Аксиньюшка	39	Аленушка	3	черная а.
628	Аксинья	36	аленький	3	амакоза
33	аксиома	3	алерт	179	Амалия
11	аксиос	35	алеть	7	Амалья
153	Аксютка	9	алеут	4	амальгамироваться
61	Аксюша	14	алеутский	30	Амальхен
513	акт	40	Алеха	18	аман
18	обвинительный акт	1754	Алеша	8	аманат
4	публичный акт	23	Алешенька	23	Амаранта
404	актер	9	алешин	8	амба
3	провинциальный а.	119	Алешка	268	амбар
17	актерка	1977	али	6	мельничный а.
15	актерский	26	али бо	8	соляной а.
6	актерство	986	аль	10	хлебный а.
14	активный	96	Алина	27	анбар
4	актовый	22	Алифан	12	амбаркадер
276	актриса	36	алкаты	6	амбарный
6	провинциальная а.	72	Алкид	11	амбарушка
7	французская а.	9	алкоголь	13	амбарчик
57	акула	26	Алла	4	амбициозно

4	амбициозный	8	аггел	50	антипатия
7	амбиционный	11	андел	27	Антипка
79	амбиция	3	ангел-барышня	5	антипод
3	анбиция	3	ангел-спаситель	7	антитеза
7	амбра	7	ангел-утешитель	211	антихрист
21	амбразура	50	ангел-хранитель	53	антихристов
7	амбре	11	ангелоподобный	56	античный
46	амвон	16	ангелочек	4	антологический
118	Амвросий	12	ангельски	351	Антон
14	Амедей	179	ангельский	46	Антонида
11	Аменаида	15	а. душа	63	Антоний
291	Америка	15	а. чин	38	Антонина
85	американец	28	ангельчик	138	Антонио
3	американизм	5	англез	10	антонов (огонь)
10	американка	8	англизированный	5	антоновский
154	американский	5	англизированной	48	Антоша
3	аметист	922	английский	102	Антошка
3	аметистовый	71	аглицкий	82	антракт
3	ами (ф)	80	Английский клуб	6	антраша
4	амигдалина	4	англиканский	5	антре
124	аминь	943	англичанин	72	антрепренер
6	амишка	6	агличанин	71	антресоль
27	амнистия	4	англичанин-управляющий	36	Антропка
15	Амос	84	англичанка	3	антрополог
22	ампула	499	Англия	3	антропологический
8	ампутация	8	англоман	3	антропофаг
3	ампутировать	9	англосаксонский	6	антропофагия
38	амуниция	5	ангорский	3	анфан (ф)
3	амуничка	24	Анджелика	37	анфилада
141	амур	38	андреевский	15	а. комнат
4	амуриться	2959	Андрей	574	Анфиса
8	амурничать	29	Андреюшка	12	анфисин
25	амурный	34	Андрьян	57	Анфуса
5	амурский	163	Андрюша	9	анчар
11	амурчик	6	андрюшин	3	анчутка
20	амфитеатр	60	Андрюшка	3	аншеф
19	амфитрион	11	аневризм	20	Анька
365	ан	441	анекдот	51	Анюта
3	анабаптист	47	анекдот о	9	анютины (глазки)
4	анадысь	4	скандальный а.	23	Анютка
104	анализ	11	анекдотец	71	Аня
7	психологический а.	7	анекдотик	4	аониды
3	химический а.	13	анекдотический	47	апартамент
50	анализировать	5	анемия	6	апатически
10	аналогия	36	Анет	19	апатический
108	аналой	45	Анета	15	апатично
50	налой	10	Анеточка	5	апатичность
76	ананас	94	Анисим	28	апатичный
7	ананасный	4	анисовка	129	апатия
5	ананасовый	4	анисовый	5	апеллировать
163	Ананий	300	Анисья	8	апелляционный
4	анархист	10	аничкин	19	апелляция
4	анархический	29	аничков	86	апельсин
28	анархия	5411	Анна	11	апельсинный
55	Анастасия	13	Анна (орден)	5	апельсинчик
80	Анатолий	46	Анненков	49	аплодировать
304	Анатолий	28	анненский	41	аплодисмент
5	анатом	11	а. лента	37	апломб
3	анатомировать	211	Аннинька	16	апогей
21	анатомический	324	Аннушка	48	Апокалипсис
31	анатомия	4	аномалия	11	апокалипсический
158	анафема	7	аноним	10	апокрифический
40	анафемский	7	анонимно	91	Аполлинария
15	Анафролия	48	анонимный	213	Аполлон
10	Анахарсис	33	а. письмо	6	аполлонов
12	анакорет	4	ансамбль	145	Аполлос
7	анахорит	29	антагонизм	6	апология
12	анахронизм	8	антагонист	15	апоплексический
7	ангажемент	8	антецедент	14	а. удар
31	ангажировать	36	антик	4	аполексия
5	ангальт-цербстский	4	антикритика	3	апорт
1220	ангел	3	антимония	153	апостол
36	а. божий	3	антиномия	6	апостольник
11	а. господень	19	Антиох	35	апостольский
225	ангел мой	200	Антип	3	апостольство
21	а. небесный	11	антипатичный	3	апостольствовать

8	апострофа	10	Арефий	23	артерия
15	апотеоза	11	Арзамас	45	артикул
4	апофеоз	4	арзамасский	88	артиллерийский
19	аппарат	9	ариаднин	64	артиллерист
303	аппетит	865	Арина	153	артиллерия
16	аппетитно	28	Аринушка	6	антиллерия
13	аппетитный	86	Аристарх	400	артист
18	Апраксин двор	10	Алистарх	14	артистически
6	апраксинцы	236	аристократ	131	артистический
379	апрель	39	аристократизм	45	артистка
17	апрельский	8	аристократик	7	артишок
7	апробировать	7	аристократически	75	Артур
175	Апроська	261	аристократический	47	арфа
23	апроськин	3	аристократично	5	элова а.
164	аптека	13	аристократишко	22	арфистка
24	аптекарский	136	аристократия	21	аржалук
35	аптекарьша	34	аристократка	11	аржалучек
126	аптекарь	3	аристократничать	80	архангел
11	старый а.	3	аристотелев	52	Архангельск
10	аптечка	3	арифметик	118	архангельский
11	аптечный	115	арифметика	6	архар
6	апшеронский	5	арихметика	4	архаровцы
5	апшеронцы	21	арифметический	7	аржей
17	араб	9	а. задача	18	археолог
17	арабеска	91	Ариша	10	археологический
31	арабский	10	Аришка	5	археология
8	аравийский	89	ария	5	архибестия
5	аравитянин	12	итальянская а.	116	архив
3	арака	74	арка	35	архивариус
222	Аракчеев	9	аркада	36	архивный
8	аракчеевский	744	Аркадий	4	архидиакон
78	арап	162	Аркадия	109	архиепископ
11	арапия	13	аркадский	4	архиепископия
17	арапка	20	аркан	5	архиепископский
58	арапник	35	Аркаша	608	архиерей
6	араповский	103	Аркашка	22	архирей
3	арапчик	3	аркебузир	143	архиерейский
21	арапчонок	4	аркольский	14	архиерейство
35	Арарат	7	арлекин	178	архимандрит
45	араратский	43	Арманс	3	архимандричий
36	арба	10	армеец	9	архимедов
26	Арбат	78	армейский	6	а. рычаг
15	арбатский	5	армиды	88	Архип
110	арбуз	847	армия	11	архипастырский
7	арбузный	12	неприятельская а.	17	архипастырь
33	аргамак	72	русская а.	17	архипелаг
5	аргонавт	35	французская а.	17	Архипушка
65	аргумент	16	целая а.	8	архистратиг
4	аргументация	14	Армфельд	3	архитектон
10	аргус	151	армяк	110	архитектор
394	Ардальон	113	армянин	52	архитектура
47	Ардалион	5	армянка	6	старинная а.
3	ардальоновский	26	армянский	25	архитектурный
12	Ардатов	5	а. церковь	5	архитриклин
5	ардатовский	3	армячина	348	аршин
96	арена	7	армячишко	4	проглотить а.
6	широкая а.	5	армяшка	9	аршинник
103	аренда	3	арнауты	26	аршинный
25	арендатор	5	арника	28	арьергард
3	арендатель	108	аромат	13	ариергард
4	арендаторша	16	ароматический	7	асаже
23	арендный	38	ароматный	60	Асаф
15	ареновать	29	арсенал	85	асессор
31	ареопаг	74	Арсений	68	коллежский а.
305	арест	10	Арское поле	5	асессорский
1090	арестант	90	Артамон	3	асессорша
5	политический а.	9	артачиться	53	аскет
22	арестантик	4	артелец	25	аскетизм
112	арестантка	303	артель	26	аскетический
194	арестантский	5	рабочая а.	230	Асклипиодот
37	а. рота	4	тюремная а.	3	асклипиодотовский
3	арестация	3	артельно	56	аспид
8	арестование	47	артельный	6	аспидный
422	арестовать	58	артельщик	3	аспидов
8	арестовывать	33	Артем	3	аспирация
47	Арефа	199	Артемий	22	ассамблея