

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

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

Том 2

Studia philologica

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ РУССКОГО ЯЗЫКА им. В. В. ВИНОГРАДОВА

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

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

Том 2



Издательский Дом ЯСК
Москва 2016

УДК 811.161.1
ББК 81.2 Рус
Ш 17



Издание подготовлено при финансовой поддержке
Российского гуманитарного научного фонда (РГНФ)
проект № 16-04-16022

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. 2. М.:
Издательский Дом ЯСК, 2016. — 848 с. — (Studia philologica.)

ISBN 978-5-9907947-7-1

В соответствии с двояким характером исследования в целом настоящий том включает две части. В части 4 дистрибутивно-статистический анализ (ДСА) проводится на материале 15 миллионов слов, непосредственно примыкающих друг к другу в текстах русской прозы 1850—1870-х гг. Тот же формальный метод прилагается к английскому корпусу (4 миллиона слов романов Диккенса). И в том, и в другом случае ДСА приводит к открытию дистрибутивных классов и некоторых грамматических конструкций. Часть 5 представляет собой частотный словарь русского корпуса, показывающий распределение слов по основным жанрам, по микрожанрам и по текстам 25 писателей.

В существенно расширенном виде том 2 представлен на компакт-диске.

УДК 811.161.1
ББК 81.2 Рус

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

ISBN 978-5-9907947-7-1

© А. Я. Шайкевич, В. М. Андриющенко, Н. А. Ребецкая, 2016
© Издательский Дом ЯСК, 2016

Часть 4

**ДИСТРИБУТИВНО-СТАТИСТИЧЕСКИЙ АНАЛИЗ
В МИНИМАЛЬНОМ ИНТЕРВАЛЕ**

Напомним основное направление дистрибутивно-статистического анализа (ДСА), сформулированное в т. 1 нашей монографии.

Ставится задача описания языка по большому корпусу текстов, в котором представлены цепочки букв, разделенные пробелами (= графические слова). Эта задача решается в несколько этапов, каждому из которых соответствует свой интервал текста, т. е. задаваемый способ членения корпуса. Первый этап анализа (микроинтервал) предполагает изучение комбинаторики букв в рамках графического слова.

На следующем этапе (в минимальном интервале) изучается комбинаторика графических слов внутри бинарных словосочетаний, т. е. пар слов, непосредственно предшествующих одно другому.

Семантическая информация в ДСА не используется.

4.1. РАБОТА ДСА В МИНИМАЛЬНОМ ИНТЕРВАЛЕ НА МАТЕРИАЛЕ КОРПУСА РУССКОЙ ПРОЗЫ

4.1.0. СИСТЕМАТИЗАЦИЯ МОРФОЛОГИЧЕСКИХ ДАННЫХ

Эффективной работе ДСА в минимальном интервале должна предшествовать окончательная систематизация морфологических данных, полученных на микроинтервале. Вернемся к результатам п.п. 2.2.4 и 2.2.5 (т. 1, с. 124—127), где приведен список 43 агрегированных парадигм с соответствующими наборами суффиксов. Этот список может быть пересмотрен с учетом двух важных процедур.

1. Следует учесть ограничения на комбинаторику букв, отмеченные в табл. 1.8 и 1.9 (т. 1, с. 47—48).

2. Следует допустить интерпретацию набора ассоциированных суффиксов как отражение нескольких парадигм, надстраиваемых одна над другой в результате деривации основ.

В качестве примера комбинаторных ограничений обратимся к букве К. Табл. 1.8 показывает, что справа от этой буквы неслучайно часто появляется И и, напротив, не встречается буква Ы¹. Теперь становится ясным, что парадигмы /17/ ТЯЖЕЛ-ый и /20/ ВЫСОК-ий² различаются ровно этим одним орфографическим ограничением. Две парадигмы можно объединить в одну.

Набор суффиксов парадигмы /19/ БЛАГОРОДНЕ-йший целиком входит в набор парадигмы /18/ ХОРОШ-ий. Анализ 39 слов с суффиксом *-йш-* покажет, что 35 из них присоединяются к основам все той же парадигмы /17/ (*благороднейший, богатейший, важнейший, вернейший, главнейший, глупейший, длиннейший, добрейший, добродушнейший, злейший, любезнейший, любимейший, малейший, милейший, нежнейший, новейший, огромнейший, отличнейший, первейший, покорнейший, полнейший, почтеннейший, превосходнейший, преданнейший, прекраснейший, прелестнейший, светлейший, сильнейший, скорейший, старейший, счастливейший, ужаснейший, умнейший, честнейший, чистейший*), слово *святейший* производно от *святой* (особого варианта парадигмы /17/), а три слова (*дальнейший, искреннейший, позднейший*) окажутся в составе парадигмы /18/. Сходным образом ведет себя суффикс *-айш* (*ближайший, величайший, высочайший, глубочайший, мельчайший, низжайший, строжайший, тончайший*) с тем отличием, однако, что сопровождается изменением конечной буквы производящей основы по правилу $k > ч, з > ж$ (и более редкого $з > ж$).

К парадигме /18/ ХОРОШИЙ присоединятся наборы глагольных суффиксов *-вшая, -вшие, -вший, -вших* (около двухсот примеров), *-ющие, -ющий, -ющим, -ющих* (103 примера), а в парадигме /17/ окажется набор глагольных суффиксов *-емое, -емые, -емый, -емых* (*делаемый, желаемый, занимаемый, именуемый, называемый, обвиняемый, обожаемый, ожидаемый, описываемый, подаваемый, получаемый, предлагаемый, предполагаемый, прерываемый, преследуемый, сдерживаемый, скрываемый, требуемый, уважаемый*). К этим 19 примерам присоединятся композиты (*многоуважаемый*) и слова с префиксом *не-* (*недосягаемый, неминуемый, необитаемый, неподражаемый, нескончаемый, неумолкаемый*).

Оставаясь в рамках микроинтервала, попробуем систематизировать морфологические результаты. Ограничим поле наблюдения двумя порогами частоты: пусть снизу это будет частота 5, а сверху — 5000. Отбросив самые редкие слова, мы застрахуемся от вторжения нерегулярных вариантов, многих имен собственных или простых опечаток (ср. *столпимши, столплением, столпления, столповиск, столповое, столповщины... яшшенка, яшшенки, яшшенку,*

¹ Точно такое же ограничение распространяется на буквы Г и Х.

² Номера парадигм из списка п. 2.2.4 будут даваться в косых скобках.

яшшенку-глазунью, яй, яйва, яйве, яйве-то, яйле, яйлу). Отбрасывание самых частых слов будет полезным при попытках суммирования каких-то морфологических показателей, где функциональные слова могли бы статистически подавить регулярные явления, наблюдаемые в массе не самых частых слов.

После введения этих двух порогов общее число разных графических слов (первоначально 458 тысяч) уменьшилось почти вдвое. Введем еще одно ограничение. Установим цифру 3 как минимальное число суффиксов, необходимое для попадания основы в пул возможных участников регулярной парадигматики. В нашем уменьшенном вдвое поле наблюдения встретилось слово *кабы* с частотой 956, но ни потенциальная основа *каб-*, ни основа *кабы* не имеет других продолжений-суффиксов из числа парадигм, открытых на микроинтервале. Основа *кавалькад-* имеет два продолжения *-а* и *-ы*, оба они входят в парадигмы /17/ и /27/, но мы пока исключаем их из рассмотрения (слова *кавалькаде* с частотой 4, *кавалькадой* и *кавалькаду* с частотой 1 заставили бы нас выбрать парадигму /27/ ПРИЧИН-а, но эти слова не преодолели частотный порог). Слово *кавардак* с частотой 8 не имеет продолжений. Но основа *кавказ* имеет 4 продолжения: *-0*, *-а*, *-ец*, *-ом*, три из которых заставляют отнести эту основу в парадигму /21/ ОБРАЗ, что хорошо соответствует ожиданиям человека, знающего русский язык. Наконец основа *кавказск-* имеет 6 продолжений: *-ие*, *-ий*, *-их*, *-ого*, *-ой*, *-ом*, все они указывают на главную адъективную парадигму. К этой же парадигме отнесем основу *кажд-* с 14 продолжениями: *-ая*, *-о*, *-ого*, *-ое*, *-ой*, *-ом*, *-ому*, *-ою*, *-у*, *-ую*, *-ые*, *-ый*, *-ым*, *-ых*. Тринадцать из них входит в набор парадигмы /17/, но появился и новый кандидат в суффиксы — продолжение *-у*. Если бы в рамках данной парадигмы *-у* появлялось многократно, пришлось бы расширить парадигму. Можно ожидать, что и перечень парадигм подвергнется пересмотру: объединятся парадигмы /5/ и /6/, /9/ и /10/, /22/ и /23/, /25/ и /26/; некоторые парадигмы исчезнут, появятся новые.

Теперь число основ, удовлетворяющих нашим требованиям, сократилось до 13 тысяч. Дадим эскиз будущего списка в новой нотации, оставляя в косых скобках старые названия и номера парадигм.

А	/Глагольные парадигмы/
Аа	/1/, /3/
Аб	/5/, /6/
Ас	/9/, /10/
Ад	/12/
В	/Субстантивные парадигмы/
Ва	/21/, /22/
Вб	/25/, /26/
Вс	/35/, /36/
Вд	/37-40/
Ве	/27/, /30/, /31/
Вф	/28/
С	/ Адъективные парадигмы /
Са	/17/, /20/
Сб	/18/, /19/

4.1.0.1. Адъективные парадигмы

Начнем наш обзор с самого ясного класса — С.

Парадигма Са (более 2500 основ) содержит самый большой набор суффиксов: *-а*, *-ая*, *-ее*, *-ей*, *-о*, *-ого*, *-ое*, *-ой*, *-ом*, *-ому*, *-ою*, *-ую*, *-ы*, *-ые*, *-ый*, *-ым*, *-ыми*, *-ых*. К этим восемнадцати суффиксам теперь добавился еще один *-0* (нулевой суффикс), наблюдаемый почти у двухсот основ. Назовем самые частые формы:

белокур	17	глух	43	женат	273	ленив	40
богат	301	гол	19	жесток	54	лют	20
бодр	47	горазд	46	жив	725	мал	131
быстр	19	горд	147	задумчив	88	мертв	61
вежлив	57	готов	1690	застенчив	36	мил	207
велик	271	груб	84	здоров	505	милосерд	22
весел	331	густ	19	здрав	22	милостив	197
виноват	1489	далек	54	зелен	22	млад	50
вспыльчив	18	дик	34	знаком	459	молод	359
всяк	218	доверчив	19	красив	117	наг	25
высок	91	един	61	ласков	121	находчив	18

невелик	32	одинок	36	справедлив	59	тяжел	73
невредим	38	подл	30	стар	310	утрюм	54
недалек	23	помешан	37	строг	109	учен	24
недвижим	22	прав (?)	982 ¹	стыдлив	19	учтив	26
нездоров	139	приветлив	26	суров	40	колост	28
незнаком	41	проклят	54	счастлив	715	храбр	23
некрасив	19	пьян	567	сыт	136	худ	85
необразован	20	расчетлив	18	таков	534	цел	114
необходим	79	ревнив	33	тверд	77	честолюбив	18
неправ	45	самолюбив	60	терпелив	20	чист	104
несчастлив	108	свят	93	тих	91	чужд	33
нетерпелив	19	скуп	65	толст	40	широк	41
обязан	504	слаб	222	трезв	36	эдак	307
одинаков	18	слеп	38	туг	20	этак	1115

В тех случаях, когда при отбрасывании положительных суффиксов исход основы содержит сочетание согласных букв, перед последней буквой появляется гласная. Так возникают закомуфлированные нулевые формы: *долот* 26 (долгий), *полон* 201 (полный), *смешон* 119 (смешной). Особенно характерны они при основах, кончающихся буквой -к.

близок	118	короток	36	ловок	79	резок	37
гадок	63	крепок	71	мяжок	32	робок	36
дерзок	48	кроток	40	неловок	32	сладок	20
жалок	77	легок	54	низок	47	тонок	23
						узок	23

Если основа заканчивается суффиксом -н, в нулевой форме перед этим суффиксом появляется -е⁻² (на -е- заменяется и мягкий знак в последовательности -льн). В нашем материале встретилось 280 форм с -ен. Приведем важнейшие:

беден	145	добродетелен	10	нежен	72	почтителен	40
безмолвен	10	добродушен	15	неизвестен	36	прекрасен	47
безобразен	11	доволен	908	ненавистен	25	прелестен	12
беспокоен	24	должен	4821	неопытен	17	привычен	11
беспощаден	10	дружен	45	неосторожен	19	приличен	18
бессилен	20	дурен	81	неподвижен	40	причастен	11
благодарен	319	жаден	20	непонятен	10	приятен	66
благоразумен	12	желчен	12	неприятен	55	простодушен	14
благороден	79	забавен	17	неравнодушен	32	противен	116
блуден	193	заметен	30	несносен	15	равен	19
блуден	565	замечателен	38	несогласен	14	равнодушен	109
важен	67	занимателен	11	несчастен	56	радостен	14
великодушен	41	звучен	10	ничтожен	41	развязен	19
верен	124	злобен	14	нужен	681	раздражителен	29
виден	185	знатен	22	опасен	48	разумен	16
виновен	82	известен	231	опрятен	12	решителен	13
властен	43	изящен	10	опытен	16	ровен	76
внимателен	39	интересен	53	оригинален	14	свободен	184
возможен	30	искусен	14	осторожен	47	серьезен	56
волен	71	исправен	17	остроумен	13	силен	215
вреден	27	красен	42	отвратителен	17	скромен	65
выгоден	17	любезен	120	откровенен	82	скрытен	11
годен	52	любопытен	27	памятен	19	скучен	57
голоден	82	мнителен	30	печален	46	славен	21
грешен	61	мрачен	83	повинен	41	слышен	222
грозен	33	надобен	37	подобен	11	смирн	42
грустен	76	наивен	21	подозрителен	19	снисходителен	30
грязен	13	небрежен	16	покорен	20	согласен	687
деликатен	26	невинен	37	понятен	36	спокоен	279
деятелен	18	недоволен	289	послушен	16	способен	383

¹ Вопросительный знак при слове *прав* указывает на большую вероятность омонимии.

² Вышеприведенный пример *смешон* — единственное исключение.

странен	65	сходен	21	удачен	12	холоден	86
страшен	142	темен	15	удобен	16	чуден	18
строен	32	тесен	12	ужасен	46	черен	13
сумрачен	17	труден	22	умен	371	честен	99
						ясен	69

Теперь, когда набор суффиксов парадигмы Са окончательно определен, можно обратиться к анализу относительной частоты отдельных суффиксов. Материал пятидесяти частых основ¹ с общей частотой 227 680 словоупотреблений дает такую картину:

-0	,046	-о	,126	-ому	,017	-ый	,117
-а	,030	-ого	,073	-ою	,010	-ым	,058
-ая	,073	-ое	,066	-ую	,038	-ыми	,028
-ее	,015	-ой	,087	-ы	,020	-ых	,066
-ей	,004	-ом	,028	-ые	,088		

В табл. 4.1 даются абсолютные частоты сочетаний суффиксов парадигмы с некоторыми характерными основами².

Таблица 4.1

	бел-	глуп-	готов-	жив-	мил-	молод-	нов-	полн-	русск-	свят-
0	26	465	1690	725	208	359	15	201	-	93
а	87	212	707	334	205	271	25	200	6	19
ая	379	311	60	285	1105	1129	772	372	487	256
ее	40	155	2	122	80	-	9	54	-	18
ей	17	27	-	80	36	1	-	4	-	13
о	34	765	494	771	418	33	137	1606	-	152
ого	331	109	46	360	176	1303	1262	312	958	291
ое	267	146	55	350	194	234	1148	617	260	136
ой	552	148	35	624	211	4191	912	471	872	536
ом	621	31	35	43	42	124	497	484	256	49
ому	69	40	6	44	56	346	290	38	220	63
ою	89	30	11	45	31	63	109	153	33	36
ую	242	107	31	182	106	338	873	357	273	132
ы	55	146	544	196	74	178	35	142	9	17
ые	682	219	89	234	258	1344	1323	192	902	220
ый	564	349	178	8	1790	-	1712	482	1158	84
ым	304	134	78	214	147	594	646	422	398	166
ыми	376	57	22	103	36	213	269	87	189	98
ых	440	104	47	398	95	1190	967	62	962	252
	5175	3555	4130	5118	5267	11911	11001	6256	6985	2631

При полной независимости вероятностей суффиксов и основ мы ожидали бы, что слово *молодой* должно появиться $(11\ 911 \times 0,117) \approx 1393$ раза, между тем оно не встретилось ни разу; математическое ожидание формы *молодой* = 1036, реально же оно появилось 4191 раз. Униформное распределение суффиксов дало бы математическое ожидание 738 у формы *полно*, реально же ее частота составила 1606.

Для оценки неслучайности отклонений реальных частот от математического ожидания снова воспользуемся нашей основной формулой (т. 1, с. 31)

$$S = (x - m - 1) / \sqrt{m}.$$

Приняв в качестве порога $S = 2$, заменим превышения над математическим ожиданием знаком плюс. Тогда из табл. 4.1 получим табл. 4.2.

¹ В эту группу входят: *нов-*, *стар-*, *добр-*, *цел-*, *молод-*, *русск-*, *велик-*, *бел-*, *мил-*, *разн-*, *полн-*, *черн-*, *бедн-*, *знаком-*, *высок-*, *готов-*, *стран-*, *собствен-*, *умн-*, *весел-*, *красн-*, *длинн-*, *сильн-*, *особенн-*, *счастлив-*, *прост-*, *родн-*, *темн-*, *жив-*, *больн-*, *подобн-*, *прекрасн-*, *богат-*, *свят-*, *глуп-*, *чист-*, *огромн-*, *пуст-*, *честн-*, *главн-*, *красив-*, *тяжел-*, *тих-*, *пьян-*, *довольн-*, *широк-*, *важн-*, *бледн-*, *благородн-*, *несчастн-*. За пределами группы оказалось 8 самых частых основ (*котор-*, *так-*, *друг-*, *как-*, *сам-*, *перв-*, *всяк-*, *кажд-*), выходящих за границы нашего поля наблюдения.

² Напомним, что орфографические *-и* после *к-*, *г-* и *х-* заменены на *-ы*.

Таблица 4.2

	бел-	глуп-	готов-	жив-	мил-	молод-	нов-	полн-	русск-	свят-
0		+	+	+						
а		+	+	+	+					
ая		+			+	+				+
ее		+		+						
ей		+		+						
о		+		+				+		
ого						+	+		+	+
ое							+	+		
ой	+					+			+	+
ом	+						+	+	+	
ому						+	+		+	+
ою	+							+		
ую							+	+	+	
ы		+	+	+						
ые	+					+	+		+	
ый					+		+		+	
ым										
ыми	+									+
ых	+					+	+		+	+

Таблицы такого типа могут дать информацию о внутренней неоднородности данных — как об основах, так и о суффиксах. Основы *глуп-* и *жив-* имеют общие плюсы в шести строках, что говорит об их сходстве. С другой стороны, суффиксы *-0*, *-а* и *-ы* имеют общие плюсы в трех столбцах (*глуп-*, *готов-* и *жив-*), что также говорит об их дистрибутивном (а может быть, и о семантическом) сходстве.

Заменив частоты на плюсы, мы превратили количественные данные в качественные и можем воспользоваться коэффициентом связи, введенным Дж. Юлом еще в 1900 г. [Upton, Cook 2014: 435]:

$$Q = (ad - bc)/(ad + bc).$$

Применим эту формулу к случаю суффиксов *-0* и *-а* в таблице, аналогичной табл. 4.2, но охватывающей весь список 50 частых основ. Оба суффикса появляются как плюсы у 9 основ (*бледн-*, *больн-*, *богат-*, *глуп-*, *готов-*, *довольн-*, *жив-*, *счастлив-*, *умн-*). Две основы дают плюс только для суффикса *-0* (*знаком*, *пьян*), четыре основы имеют плюс только у суффикса *-а* (*велика*, *добра*, *мила*, *чиста*). Эти данные представлены в четырехклеточной таблице:

	-0	не-0	Итого	
-а		9 (a)	2 (b)	11
не-а		4 (c)	35 (d)	39
Итого	13		37	50

Отсюда $Q = (9 \times 35 - 2 \times 4) / (9 \times 35 + 2 \times 4) = 307/323 = 0,95$ — необычайно высокий показатель дистрибутивного сходства. Для пары суффиксов *-0* и *-ы* $Q = 0,99$, для *-а* и *-ы* $Q = 0,75$. Все это дает нам некоторый кластер из трех суффиксов в рамках общей адъективной парадигмы. Содержательная оценка этого кластера сразу бы определила — речь идет о кратких прилагательных; быть может, минимальный интервал обнаружит здесь что-нибудь средствами ДСА. Как бы то ни было, существование кластера есть доказанный факт. Ниже будут показаны некоторые черты внутренней структуры субстантивных парадигм, открываемые на микроинтервале.

Обратимся теперь к неоднородности основ.

Статистически значимое сокращение форм с суффиксом *-ый* (как в случаях *живый* и *святой*) или их полное отсутствие не может быть объяснено в рамках ДСА, ведь ударение почти не находит отражения в русской орфографии. Подобное поведение обнаружено у 146 основ (или 6 %), ср. *босой*, *восьмой*, *второй*, *иной*, *кривой*, *крутой*, *лихой*, *молодой*, *немой*, *плохой*, *простой*, *прямой*, *пустой*, *рябой*, *святой*, *седой*, *седьмой*, *скупой*, *сухой*, *тугой*, *тупой*, *холостой*, *шестой* и т. д. В основах с суффиксом *-ов-* подобные случаи составляют до 40 % (*боковой*, *вековой*, *верховой*, *вестовой*, *годовой*, *городовой*, *громовой*, *деловой*, *долговой*, *круговой*, *мастеровой*, *меховой*, *мировой*, *моровой*, *мостовой*, *носовой*, *передовой*, *полковой*, *половой*, *родовой*, *роковой*, *рядовой*, *становой*, *трудовой*, *часовой* и др.). В словах с суффиксом *-н-* доля их меньше 10 %, но представлены они очень широко (*головной*, *грудной*, *дверной*, *двойной*, *земной*, *крепостной*, *лесной*, *мясной*, *ночной*, *основной*, *отставной*, *переводной*, *перекладной*, *пивной*, *портной*, *потайной*, *резной*, *речной*, *ржаной*, *родной*, *ручной*, *смешной*, *стальной*, *стенной*, *степной*, *тройной*, *хмельной*, *цветной*, *цепной*, *четвертной*, *шерстяной* и т. п.). Встречаются они и среди основ

с суффиксом *-ск-* (4 %), ср. *воровской, городской, донской, людской, мастерской, мирской, морской, мужской, слободской, тверской, ямской*.

Рассмотренная парадигма *Са* включает огромное число основ с суффиксами. Самый частый суффикс *-н-* образуется по преимуществу адъективы от субстантивов (более 80 %). При этом происходит изменение конечной буквы производящей основы:

-к > -ч,
-ц > -ч,
-г > -ж,
-х > -ш.

Назовем эту схему основным чередованием согласных (она будет неоднократно повторяться у суффиксов именных парадигм). Примерами суффикса *-н-* могут служить:

брачн-	(брак)	ручн-	(рука)	бумажн-	(бумага)
вечн-	(век)	сказочн-	(сказка)	денежн-	(денег)
восточн-	(восток)	табачн-	(табак)	дорожн-	(дорога)
горячечн-	(горячка)	фабричн-	(фабрика)	дружн-	(друг)
желудочн-	(желудок)	цветочн-	(цветок)	жемчужн-	(жемчуг)
загадочн-	(загадка)	циничн-	(циник)	каторжн-	(каторга)
звучн-	(звук)	чахоточн-	(чахотка)	книжн-	(книга)
копеечн-	(копейка)	шуточн-	(шутка)	ночлежн-	(ночлег)
крошечн-	(крошка)	ярмарочн-	(ярмарка)	острожн-	(острог)
кулачн-	(кулак)			отважн-	(отвага)
лихорадочн-	(лихорадка)	больничн-	(больница)	пирожн-	(пирог)
молочн-	(молоко)	горничн-	(горница)	присяжн-	(присяга)
научн-	(наука)	единичн-	(единица)	радужн-	(радуга)
оброчн-	(оброк)	заграничн-	(граница)	снежн-	(снег)
отрывочн-	(отрывок)	конечн-	(конец)	судорожн-	(судорога)
ошибочн-	(ошибка)	мельничн-	(мельница)	тревожн-	(тревога)
политичн-	(политика)	месячн-	(месяц)	южн-	(юг)
порочн-	(порок)	пшеничн-	(пшеница)		
порядочн-	(порядок)	сердечн-	(сердце)	воздушн-	(воздух)
праздничн-	(праздник)	солнечн-	(солнце)	грешн-	(грех)
практичн-	(практика)	столичн-	(столица)	душн-	(дух)
публичн-	(публика)	уличн-	(улица)	потешн-	(потеха)
пушечн-	(пушка)			смешн-	(смех)
речн-	(река)			страшн-	(страх)
				успешн-	(успех)

Как правило, производная основа имеет меньшую частоту, чем производящая основа. Обратное соотношение в парах *мрачн-* (*мрак*), *скучн-* (*скука*), *должн-* (*долг*), *нежн-* (*нега*) заставляет предполагать ослабление семантической зависимости производной основы от производящей. В некоторых случаях основа с исходом на *-чн-* не укладывается в основную схему чередований, и тогда можно предположить существование более протяженного суффикса *-ичн-* (в парах *будничн-* — *будни*, *вторичн-* — *второй*, *симпатичн-* — *симпатия*, *энергичн-* — *энергия*). Сходная трактовка напрашивается при анализе основ *зажиточн-* и *распивочн-*.

Можно считать доказанным существование отглагольного суффикса *-тельн-*, подтверждаемого внушительным списком:

блистательн-	изумительн-	обстоятельн-	отрицательн-
взыскательн-	исключительн-	обходительн-	очаровательн-
влиятельн-	ласкательн-	обязательн-	ощутительн-
внимательн-	медлительн-	оглушительн-	повелительн-
внушительн-	мнительн-	одобрительн-	подозрительн-
возмутительн-	мстительн-	окончательн-	позволительн-
вопросительн-	мучительн-	омерзительн-	положительн-
выразительн-	назидательн-	оправдательн-	поразительн-
действительн-	наставительн-	оскорбительн-	последовательн-
заключительн-	настоятельн-	ослепительн-	поучительн-
замечательн-	нюхательн-	основательн-	почтительн-
занимательн-	обаятельн-	осязательн-	предварительн-
заразительн-	обличительн-	отвратительн-	представительн-
значительн-	обольстительн-	отличительн-	предупредительн-

предусмотрительн-	решительн-	стеснительн-	удовлетворительн-
презрительн-	ругательн-	стремительн-	указательн-
привлекательн-	снисходительн-	томительн-	умилительн-
пригласительн-	соблазнительн-	требовательн-	уморительн-
приготовительн-	созерцательн-	трогательн-	унизительн-
признательн-	сознательн-	убедительн-	успокоительн-
продолжительн-	сомнительн-	уважительн-	утвердительн-
проницательн-	сострадательн-	увеселительн-	утешительн-
раздражительн-	спасительн-	увлекательн-	утомительн-
рассудительн-	старательн-	удивительн-	чувствительн-

Популярность суффикса *-н-* способствует его экспансии, в частности, в сферу европейских заимствований, ср. *гениальн-*, *колоссальн-*, *музыкальн-*, *натуральн-*, *национальн-*, *нормальн-*, *официальн-*, *патриархальн-*, *радикальн-*, *рациональн-*, *сентиментальн-*, *социальн-*, *танцевальн-*, *театральн-*, *центральнон-*. Загадочным кажется пространство *-альн-* на русские основы: *буквальн-*, *погребальн-*, *поминальн-*, *процальн-*.

Около 300 основ рассматриваемой парадигмы содержат суффикс *-ск-*, присоединяемый к субстантивным основам (суффиксация соответствует схеме В --> Са). Этот суффикс обычно вызывает основное чередование согласных (ср., однако, *петербургский*, *оренбургский*). Присоединение его к производящим основам на *-л-* сопровождается появлением *-ь-*, (*генеральский*, *дьявольский*, *сельский*, *тульский*, *уральский*). Напротив, всякий след «мягкости» может исчезнуть после *-н-* и *-р-*: *деревенский*, *конский*, *богатырский*, *зверский*, *морской*, *сибирский*, *ухарский*, *царский*. Перед данным суффиксом появляется *-е-* после конечных *-ч-* или *-ш-* производящей основы (независимо от причины появления этих букв): *всяческ-*, *греческ-*, *девическ-*, *иноческ-*, *купческ-*, *отческ-*, *творческ-*, *человеческ-*, *языческ-*; *монашеск-*, *юношеск-*. После *-ж-* твердое правило установить не удастся, ср. *дружеск-*, *княжеск-*, *пажеск-*, *супружеск-*, но *волжск-*, *калужск-*, *коллежск-*, *мужск-*, *парижск-*.

Более 60 основ кончаются на последовательность *-ическ-*, лишь одна из них (*девическ-*) целиком состоит из русских морфем; все остальные — результат адаптации европейских прилагательных (во французском на *-ique*): *гармоническ-*, *географическ-*, *героическ-*, *гимназическ-*, *готическ-*, *демократическ-*, *деспотическ-*, *дипломатическ-*, *ироническ-*, *истерическ-*, *историческ-*, *католическ-*, *классическ-*, *комическ-*, *критическ-*, *лирическ-*, *логическ-*, *магическ-*, *математическ-*, *меланхолическ-*, *механическ-*, *мистическ-*, *нервическ-*, *патетическ-*, *патриотическ-*, *педагогическ-*, *периодическ-*, *политическ-*, *поэтическ-*, *практическ-*, *психическ-*, *психологическ-*, *романическ-*, *романтическ-*, *саркастическ-*, *сатирическ-*, *симпатическ-*, *систематическ-*, *социалистическ-*, *статистическ-*, *студенческ-*, *сценическ-*, *теоретическ-*, *трагическ-*, *тропическ-*, *фанатическ-*, *характеристическ-*, *фактическ-*, *фантастическ-*, *физиологическ-*, *физическ-*, *фотографическ-*, *химическ-*, *эгоистическ-*, *экономическ-*, *эксцентрическ-*, *энергическ-*, *эстетическ-*, *юмористическ-*, *юридическ-*.

Основы на *-ск-* и *-ическ-* обладают рядом дистрибутивных ограничений — они не сочетаются с *-л-*, *-а-*, *-о-*. С другой стороны, формы с суффиксом *-и-*, по-видимому, ассоциированы не с *-л-* и *-а-*, но функционально аналогичны формам с конечным *-о* остальных основ нашей парадигмы. Так это, или не так, покажет минимальный интервал. Ср. *воровски*, *гармонически*, *детски*, *иронически*, *мастерски*, *плутовски*, *предательски*, *старчески*, *фактически*, *физически*, *философски*, *человечески*, *юношески* (всего 55 форм).

Из остальных основообразующих суффиксов упомянем:

- ист- (30 основ): *болотистый*, *волнистый*, *глинистый*, *голосистый*, *душистый*, *жилистый*, *золотистый*, *извилистый*, *каменистый*, *лесистый*, *лучистый*, *морщинистый*, *мускулистый*, *мясистый*, *неказистый*, *объемистый*, *окладистый*, *осанистый*, *отрывистый*, *порывистый*, *прерывистый*, *приземистый*, *пушистый*, *размашистый*, *речистый*, *росистый*, *серебристый*, *скалистый*, *смолистый*, *тенистый*;
- лив- (51 основа): *болтлив-*, *боязлив-*, *брезглив-*, *визглив-*, *говорлив-*, *горделив-*, *добычлив-*, *догадлив-*, *дождлив-*, *жалостлив-*, *заботлив-*, *завистлив-*, *затейлив-*, *кокетлив-*, *криклив-*, *молчалив-*, *назойлив-*, *насмешлив-*, *особлив-*, *отчетлив-*, *приветлив-*, *прихотлив-*, *причудлив-*, *прозорлив-*, *пуглив-*, *пытлив-*, *расчетлив-*, *слезлив-*, *словоохотлив-*, *смазлив-*, *сметлив-*, *смешлив-*, *сонлив-*, *справедлив-*, *стыдлив-*, *суетлив-*, *счастлив-*, *талантлив-*, *терпелив-*, *тороплив-*, *тосклив-*, *труслив-*, *удушлив-*, *уродлив-*, *услужлив-*, *хвастлив-*, *хлопотлив-*, *шутлив-*, *щекотлив-*;
- чив- (16 основ, в конечном счете, глагольного происхождения): *вкрадчив-*, *вспыльчив-*, *доверчив-*, *задумчив-*, *заманчив-*, *заносчив-*, *застенчив-*, *навязчив-*, *настойчив-*, *находчив-*, *обидчив-*, *обманчив-*, *разборчив-*, *разговорчив-*, *сбивчив-*, *уклончив-*;
- оват- (16 основ, чаще производных от адъективных же основ): *белесоват-*, *беловат-*, *глуповат-*, *голубоват-*, *грубоват-*, *желтоват-*, *зеленоват-*, *красноват-*, *плутоват-*, *подслеповат-*, *продолговат-*, *простоват-*, *сероват-*, *сутуловат-*, *угловат-*, *щеголеват-*.

Отглагольные адъективы будут рассмотрены в связи с морфологией глагола.

Сб («Мягкая парадигма адъективов»). Данная парадигма отличается от только что рассмотренной тем, что 1) суффиксы *-ого, -ое, -ой, -ом, -ому, -ою* заменяются на *-его, -ее, -ей, -ем, -ему, -ею*; 2) суффиксы *-ие, -ий, -им, -ими, -их* появляются не только после *ш-, ж-, ч-* и *щ-* (где буква *-ы* невозможна), но и в остальных случаях; 3) после указанных согласных сохраняются орфографические *-ая* и *-ую*, в остальных же случаях вместо них появляются *-яя* и *-юю*; 4) в первоначальном наборе суффиксов нет *-о*. Проверка всех основ на возможность включения в данную парадигму добавила несколько суффиксов: *-л, -а, -е, -и*. К парадигме несомненно относится 186 основ (с исходом на *ш-* — 9 основ, с исходом на *щ-* — 8 основ, с исходом на *ж-* — 18 основ, с исходом на *ч-* — 31 основа, с исходом на *н-* — 120 основ¹).

Как и в случае парадигмы Са, подсчитаем долю каждого суффикса на основании группы 50 самых частых основ с общей частотой 61 150².

Особой парадигмой Сс следует признать набор суффиксов *-е, -его, -ей, -ем, -ему, -ею, -и, -им, -ими, -их, -ю, -я*, присоединяемых к основам с исходом на *-ь*. Теряя конечный *-ь*, эти основы соединяются с суффиксом *-ий*. Группа включает 38 основ с общей частотой 9370: *бабь-, барань-, беличь-, божь-, вдовь-, воловь-, волчь-, вражь-, девичь-, заячь-, извозчичь-, казачь-, козь-, коровь-, кошачь-, куличь-, лись-, медвежь-, мужичь-, олень-, охотничь-, павлинь-, полковничь-, помещичь-, птичь-, разбойничь-, раскольничь-, ребячь-, рыбаць-, рыбь-, собачь-, собопь-, старушечь-, стерляжь-, телячь-, треть-, человекь-, чиновничь-*. За единственным исключением (*треть-*) все эти основы производны от субстантивных основ. К основному чередованию согласных в этих случаях присоединяется чередование *д > ж* (*медвежь-, стерляжь-*) и *т > ч* (*ребячь-, телячь-*)

Роль отдельных суффиксов в адъективных парадигмах выглядит следующим образом:

	Са	Сб-ш, ж, ч, щ	Сб-н	Сс
-о	4,6%	5,1%	...	-
-а	3,0%	4,5%	0,1%	-
-ая	7,3%	6,0%	-яя 5,4%	-я 9,4%
-ую	3,8%	3,8%	-юю 5,8%	-ю 7,4%
-о	12,6%			-
		-е 1,9%	1,9%	-
-ее	1,5%			
-ей	0,4%			
-ого	7,3%	-его 10,3%	10,0%	12,2%
-ое	6,6%	-ее 9,6%	10,3%	-е 6,3%
-ой	8,7%	-ей 5,6%	20,1%	10,5%
-ом	2,8%	-ем 2,8%	4,4%	5,1%
-ому	1,7%	-ему 2,7%	2,1%	2,0%
-ою	1,0%	-ею 0,7%	0,8%	0,5%
-ы	2,0%	-и 1,5%	0,1%	-
-ые	8,8%	-ие 10,4%	10,8%	-и 8,3%
-ый	11,7%	-ий 14,8%	12,8%	27,8%
-ым	5,8%	-им 9,0%	4,5%	4,2%
-ыми	2,8%	-ими 2,9%	1,3%	1,5%
-ых	6,6%	-их 8,3%	9,2%	4,8%

Как видим, в парадигме Сб основы с исходом на *-н* не присоединяют суффиксы *-л, -а, -и*³; ту же особенность наблюдаем в парадигме Сс. Конечно, приравнивание суффиксов в адъективных парадигмах несколько опережает ход алгоритма. Минимальный интервал должен подтвердить или опровергнуть такое приравнивание.

¹ Лишь у слова *синий* конечному *-н*- основы предшествует гласная. Во всех остальных случаях здесь находим либо сочетание согласный + *н*: *ближн-, нынешн-,* либо удвоенное *-нн-*: *ранн-, осенн-*.

² В эту группу входят *городнич-, дюж-, жгуч-, летуч-, леи-, лучи-, млади-, могуч-, неуклюж-, нехорош-, ниц-, общ-, падуч-, певч-, подьяч-, похож-, приезж-, прихож-, прохож-, проч-, рабоч-, рыж-, свеж-, стряпч-, сумасшедш-, суц-, тощ-, хорош-; ближн-, верхн-, весенн-, внутренн-, дальн-, домашн-, древн-, завтрашн-, замужен-, здешн-, ихн-, крайн-, летн-, лишн-, нижн-, осенн-, передн-, последн-, презж-, син-, соседн-*.

³ Феноменальная доля суффикса *-ей* частично объясняется формой *крайней*, входящей в трехсловное сочетание *по крайней мере*.

4.1.0.2. Субстантивные парадигмы

Субстантивные парадигмы объединяют тысячи основ и весьма разнообразны по своим морфологическим свойствам. Начнем с самой мощной парадигмы (Ва), включающей 2357 основ с изначально сформированным списком суффиксов: *-0, -а, -ам, -ами, -ах, -е, -ов, -ом, -у, -ы*. Табл. 141 частых основ (с общей частотой 388 тысяч словоупотреблений) дает следующие вероятности суффиксов:

-0	,406	-ам	,013	-ах	,017	-ов	,042	-у	,084
-а	,187	-ами	,027	-е	,056	-ом	,102	-ы	,057

Поступим так же, как с материалами табл. 4.2 в парадигме Са.

Таблица 4.3

Матрица коэффициентов связи (Q) между суффиксами парадигмы Ва

	-0	-ам	-ами	-ах	-ов	-ы	-а	-е	-ом	-у
-0										
-ам			88	88	68	70				
-ами				38	63	91				
-ах					52	54				49
-ов						89				
-ы										
-а								33		77
-е									44	36
-ом										60
-у										

Внутренняя структура парадигмы выступает здесь очень рельефно: несомненно доказан кластер пяти суффиксов *-ам, -ами, -ах, -ов, -ы* с характерными основами *++арестант, волос, грех, закон, мужик, офицер, охотник, помещик, поступок, праздник, преступник, работник, разговор, рассказ, солдат, чиновник*. Если по группе в целом эти пять суффиксов составляют 15,6 % общей частоты, то у 16 только что приведенных основ их доля поднимается до 45 %. В группе *++кандалы, миазмы, пожитки, тиски, финансы, штаны* только эти суффиксы и засвидетельствованы. С позиций семантической лингвистики этот кластер объясняется очень просто: соответствующие существительные часто (или даже всегда) употребляются во множественном числе.

Не столь ясен четырехчленный кластер *-а, -е, -ом, -у* с характерными основами *ветр-, город-, мир-, порядк-, рт-, снег-, стол-, столик-, страх-, ужас-, ум-*.

В рамках парадигмы обнаружено более 170 основ с исходом на *-ник-* (*бумажник, воротник, вторник, гривенник, дворник, дневник, защитник, исправник, любовник, мельник, мошенник, наследник, охотник, племянник, покойник, полковник, полтинник, помощник, понедельник, посредник, праздник, преступник, работник, разбойник, родственник, священник, советник, спутник, чайник, чиновник* и т. д.). Примерами основного чередования согласных служат *баשמачник, будочник, взяточник, источник, лавочник, лодочник; станичник; безбожник, должник, каторжник, мятежник, сапожник; грешник, орешник*. Регулярно появление *-ь-* после *-л-* производящей основы — *бездельник, начальник, отишельник, подзатыльник, пчельник, раскольник, сокольник, стольник, треугольник, умывальник, целовальник, школьник*. Семантический подход признал бы этот *-ник-* суффиксом, но зафиксировал бы несколько ошибок (*ботаник, веник, пикник, пряник, хроник, циник*).

Как и в парадигме Са, особого замечания требует суффикс *-0*. Стечение согласных букв в исходе нулевых форм преодолевается вставкой гласной перед конечным согласным: *-ол (посол, угол, хохол, чехол), -ел (вымысел, дятел, замысел, котел, орел, осел, павел, пепел, помысел, узел, умысел), -ен (молебен), -ет (египет, хребет), -ем (заем, наем), -ом (псалом), -ор (багор, бугор, свекор), -ер (ветер, костер), -ес (овес), -от (рот)*.

В максимальной степени такие нулевые форм наблюдаются у основ с конечным *-к-* (более 150 основ), ср.

беспорядок	желудок	листок	остаток	поединок
венюк	замок	мешок	оттенок	полушубок
ветерок	затылок	молоток	переулок	поросенок
висок	звонок	недостаток	песок	порядок
возок	крючок	новичок	платок	поступок
волосок	кусок	носок	подарок	потолок
десяток	лесок	образок	подбородок	припадок
рассудок	рынок	сынок	убыток	цветок
ребенок	сверток	теленок	уголок	чулок
рисунюк	список	толчок	узелок	щенок

василек	излишек	кулек	пенек	уголек
денек	конек	муженек	пузырек	флигелек
зверек	кошелек	огонек	ручеек	хорек

Само собой разумеется, основообразующий суффикс *-к-* вызывает основное чередование в производящей основе:

бычок	(бык)	пучок	(пук)	овражек	(овраг)
волчок	(волк)	ребеночек	(ребенок)	пирожок	(пирог)
дурачок	(дурак)	старичок	(старик)	прыжок	(прыг)
клочок	(клок)	сундучок	(сундук)	рожок	(рог)
куличок	(кулик)	сыночек	(сын)	сапжок	(сапог)
кусочек	(кусок)	цветочек	(цветок)	снежок	(снег)
листочек	(листок)	человечек	(человек)	петушок	(петух)
мешочек	(мешок)	червячок	(червяк)	смешок	(смех)
мужичок	(мужик)	чулочек	(чулок)	стишок	(стих)
песочек	(песок)	ящичек	(ящик)		
платочек	(платок)				

Чередование отмечено и при суффиксе *-ик* (*кончик, огурчик, пальчик*). Всего в поле наблюдения зафиксировано 43 основы с этим суффиксом: *бантик, билетик, братик, волосик, годик, гробик, грошик, дворик, дождик, домик, калачик, карлик, коврик, коник, костик, крестик, кролик, кустик, листик, мазурик, материк, матросик, мужик, носик, прутик, ротик, рублик, рыжик, садик, солдатик, старик, столбик, столик, фонарик, хвостик, часик, шарик*. С содержательной точки зрения следовало бы считать омонимом суффикс *-ик* в конце основ *историк-, классик-* и *фанатик-*, пока эти случаи остаются в общем списке.

Из других основообразующих суффиксов отметим:

- чик* (31 основа): *бубенчик, буфетчик, возчик, голубчик, графинчик, диванчик, доносчик, заводчик, закладчик, извозчик, коридорчик, лазутчик, мальчик, молодчик, обидчик, образчик, ответчик, переводчик, перевозчик, подрядчик, попутчик, приказчик, разносчик, рассказчик, рябчик, рядчик, советчик, стаканчик, челчик, чуланчик, шкапчик;*
- щик* (27 основ): *артельщик, барабанщик, бунтовщик, гробовщик, гуртовщик, денщик, заговорщик, зачинщик, каменщик, караульщик, конторщик, миллионщик, могильщик, надсмотрщик, наемщик, носильщик, обманщик, откупщик, покупатель, помещик, прапорщик, приживальщик, ростовщик, спорщик, сыщик, ходебщик, ящик.*

Из европейских субстантивных суффиксов в нашем материале представлены

- ист*: *артиллерист, артист, аферист, бонапартист, гимназист, журналист, кавалерист, кантонист, канцелярист, капиталист, колонист, коммунист, контрабандист, материалист, машинист, моралист, прогрессист, публицист, романист, семинарист, специалист, турист, фельетонист, финансист* и
- изм*: *афоризм, деспотизм, либерализм, магнетизм, механизм, мистицизм, организм, патриотизм, ревматизм, романтизм, скептицизм, спиритизм, фанатизм, цинизм.*

В пяти случаях засвидетельствованы пары с обоими суффиксами: *атеизм — атеист, идеализм — идеалист, нигилизм — нигилист, социализм — социалист, эгоизм — эгоист.*

Особого замечания требуют основы с исходом на *-ц-*. Основы *беглец, жрец, кузнец, льстец, мертвец, мудрец, подлец, чернец, чтец* вполне укладываются в парадигму Ва. У 30 основ своеобразие отмечено в нулевой форме, где перед конечной *-ц* появляется буква *-е-*: *боец, борец, вдовец, венец, глупец, гонец, гребец, дворец, жеребец, жилец, конец, купец, леденец, ловец, образец, огурец, отец, певец, писаец, пловец, продавец, птенец, рубец, самец, скворец, слепец, сорванец, стрелец, творец, чепец*. Основа *молодец-* присоединяет суффиксы либо по этом образцу, либо по нижеследующему.

Громадное большинство (более 100) остальных основ с исходом на *-ц-* к этому добавляют мену суффиксов *-ов* на *-ев* и *-ом* на *-ем*¹: *американец, братец, вольнодумец, глуповец, иностранец, кормилец, красавец, летописец, любимец, мерзавец, мизинец, младенец, незнакомец, немец, палец, полководец, республиканец, румянец, самозванец, старец, счастливец, танец* и др.

¹ У нулевых форм *месяц* и *принц*, а также у имен собственных (*аустерлиц, голыц, штольц, шульц*) в нулевой форме изменений не происходит. У основы *зайц-* в нулевой форме находим *заяц*.

Обозначим этот последний вариант парадигмы символом Ва-ц.

Еще один вариант парадигмы Ва включает 100 основ с исходом на гласную букву. В нулевой форме у них появляется *-й*, остальные суффиксы сходны с суффиксами парадигмы Вb: *-е, -ев, -ем, -и, -ю, -я, -ям, -ями, -ях*. К самым частым основам относятся: *случай, чай, край, николай, обычай, сарай, дунай; алексей, андрей, лакей, сергей, злодей, архиерей, еврей; василий, дмитрий, григорий, афанасий, порфирий, гений; покой, герой, бой, вой, обои, рой; поцелуй; негодяй*. Этот вариант парадигмы можно обозначить Ва-й.

Отличие парадигмы Вb от парадигмы Ва во многом повторяет то, что отличало парадигму Сb от парадигмы Са. В «мягких парадигмах» представлены два подтипа: основы с исходом *-ч, -ш, -ж, -щ* («шипящие») и основы с другим исходом. В обоих подтипах парадигмы Вb появляется новый суффикс *-ей, -ы* в суффиксе меняется на *-и*. Во втором подтипе гласные суффиксов последовательно меняются — *-а* на *-я* (*-я, -ям, -ями, -ях*); *-у* на *-ю*; *-о* на *-ь*; *-ом* на *-ем*. В первом подтипе частично наблюдается только последняя мутация (будем называть его Вb-чшжщ). Примерами могут служить:

-0 -а -е -ом -у -ей -ам -ами -ах
 ключ, луч, врач, меч, богач, калач, кирпич, паралич;
 грош, барыш, карандаш, торгаш;
 нож, грабеж, этаж, багаж, паж, рубеж; плащ, лещ, плющ

-0 -а -е -ем -у -ей -ам -ами -ах
 обруч, греч; марш, выигрыш, пунш;
 сторож, париж, экипаж, саквояж; товарищ, овощ

-ь -я -е -ем -ю -ей -ям -ями -ях
 голубь; червь; дождь, медведь; витязь, язь;
 корабль, король, приятель, учитель, писатель;
 конь, огонь, олень, окунь;
 государь, царь, монастырь, зверь, секретарь;
 гусь; гость.

Схема соответствий парадигм Ва и Вb оставляет не у дел суффикс *-ов*. Можно предположить, что новый член набора — суффикс *-ей* и есть функциональный коррелят суффикса *-ов* из парадигмы Ва. Аргументом в пользу такой трактовки могут быть значения Q пяти суффиксов из парадигмы Вb, определенные на группе из 40 самых частых основ:

	-ей	-и	-ям	-ями	-ях
-ей		69			68
-и			67	98	53
-ям				67	

Эти связи несколько напоминают то, что представлено в табл. 4.3, хотя со статистической точки зрения данные цифры не столь убедительны.

В нашем поле наблюдений оказалось около 200 основ, присоединяющих суффиксы по этой схеме. Некоторые очень частые основы демонстрируют комбинаторику основ и суффиксов, указывающую на какую-то промежуточную зону между Ва и Вb; примерами служат следующие формы (даны с частотами)¹:

брат	брата	брате	братом	брату					
4721	1892	116	530	782					
			братьев	братья	братьям	братьями	братьях		
			332	363	67	95	11		
друг	друга	друге	другом(?)	другу	други				
7449	2879	94	2083	1300	78				
			друзей	друзья	друзьям	друзьями	друзьях		
			540	691	106	281	25		
зять	зятя	зяте	зятем	зятю					
121	130	5	48	48					
			зятьев	зятья	зятьям	зятьями	зятьях		
			13	30	2	9	2		

¹ Включены и те формы, которые выходят за пределы поля наблюдения.

лист	листа	листе	листом	листов	листы	листам	листами	листах	
465	115	61	51	91	133	9	26	25	
				листьев	листья	листьям	листьями	листьях	
				121	231	13	94	29	
сын	сына	сыне	сыном	сыну	сынов	сыны	сынам	сынами	сынах
2363	2171	155	530	648	42	45	6	7	-
				сыновой	сыновья	сыновьям	сыновьями	сыновьях	
				174	97	26	40	9	
сосед	соседа	соседе	соседом	соседу			соседами	соседах	
309	229	11	48	114			1	1	
				соседей	соседи	соседям	соседями	соседях	
				239	220	77	64	10	

По две основы имеют слова, явно относящиеся к парадигме Вb:

князь	князя	князе	князем	князю	князей	князи			
9134	3151	150	591	1298	145	17			
				князьев	князья	князьям	князьями	князьях	
				2	108	20	27	6	
муж	мужа	муже	мужем	мужу	мужей	мужи	мужам	мужами	мужах
2443	2793	170	971	1010	144	22	3	4	1
				мужев	мужья	мужьям	мужьями	мужьях	
				5	129	32	50	10	

Отличительная изначальная черта парадигмы Вc — присутствие суффикса *-o* (*слово, место, лицо, утро, письмо, окно, облако, чувство, средство, семейство* и т. п., более 300 основ). Набор суффиксов мало отличается от парадигмы Вa, но различия в частоте суффиксов могут быть очень значительны. Сравним долю суффиксов в двух парадигмах¹:

	-0	-а	-ам	-ами	-ах	-е	-о	-ов	-ом	-у	-ы
Вa	,406	,187	,013	,027	,017	,056	-	,042	,102	,084	,057
Вc	,067	,213	,020	,031	,035	,084	,381	,003	,114	,045	,006

Суффикс *-o* занимает первое место в парадигме Вc, можно заподозрить его функциональное сходство с суффиксом *-o* в парадигме Вa. Поражает уменьшение доли *-ов* и *-ы*. Почти половина случаев *-ов* приходится на три основы — *облак-*, *яблоч-* и *мест-*. Суффикс *-ы* (орфографически *-и*) регулярно появляется после конечного *к-* основы: *личик-*, *местечк-*, *окошк-*, *словечк-*, *стеклышк-*, *ушк-*, *яблоч-*, а также у основы *плеч-*.

О функциональной роли суффикса *-o* в парадигме Вc можно судить по его коэффициентам связи с другими суффиксами

	0	а	ам	ами	ах
0		36	76	80	73
а			42	68	49
ам				63	90
ами					72
ах					

Говоря содержательно, суффикс *-o* входит в кластер суффиксов множественного числа, в какой-то мере это касается и суффикса *-а*. Появление гласных в нулевых формах повторяет схему парадигмы Вa, ср. *местечек, окошек, перышек, яичек; кресел, ремесел, чисел; стекол, тягол; писем; бревен, зерен, полотен, пятен; сукон; ведер, ребер, ядер; крылец*. Уникальна нулевая форма *яиц* у основы *яйц-*. Специфика некоторых основ с исходом *ц-* состоит (так же, как в Вa) в замене *-o* на *-е* в суффиксах: *сердце, солнце, платьице, полотенце*.

Подтипы парадигмы Вd устроены так же, как в парадигме Вb — 22 основы с суффиксом *-иц-* (*винице, голенице, детице, жилице, зрелице, кладбице, пастбице, побоице, пожарице, полчице, поприце, пристанице, прозвице, сборице, сокровице, страшилице, сходбице, туловице, убежице, удилице, училице, чудовице*); слова *море* и *поле* с суффиксами *-е, -ей, -ем, -ю, -я, -ям, -ями, -ях* и почти 700 основ с исходом на *и-* (преимущественно

¹ Для парадигмы Вc подсчет произведен по 80 самым частым основам с общей частотой 120 тысяч.

отглагольные субстантивы) и таким набором суффиксов: *-е, -ем, -и, -й, -ю, -я, -ям, -ями, -ях*. Суффиксы двух парадигм соотносятся следующим образом:

Вс	Вd	Вс	Вd	Вс	Вd
-о	-е	-ом	-ем	-ам	-ям
-е	-и	-у	-ю	-ами	-ями
-0	-й	-а	-я	-ах	-ях.

К числу самых частых основ принадлежат: *влияние, волнение, воображение, воспоминание, впечатление, выражение, движение, действие, дыхание, желание, заключение, значение, имение, мнение, молчание, название, наслаждение, несчастье, объяснение, одолжение, отсутствие, отчаяние, положение, понятие, предложение, преступление, приказание, присутствие, продолжение, расположение, решение, самолюбие, свидание, согласие, содержание, сознание, состояние, спокойствие, существование, течение, уважение, удовольствие, участие, чтение, явление*.

Более 100 основ предстают в двух вариантах: наряду с исходом на *и-* они имеют исход на *ь-*. Оба варианта не различаются в нулевой форме, где находим *-ий*. В рамках ДСА невозможно определить, есть ли семантическая разница между этими вариантами: *мгновение* и *мгновенье*, *счастье* и *счастье*, *учение* и *ученье*. Нередко один из вариантов встречается чаще, чем другой.

веселье	199	-ие 20	внимание	1540	-ье 18
воскресенье	476	-ие 29	воспитание	509	-ье 10
жалованье	494	-ие 8	занятие	317	-ье 2
житье	512	-ие 90	намерение	475	-ье 20
прощанье	220	-ие 60	пение	309	-ье 22

Форма на *-ье* может не иметь коррелята на *-ие*, ср. *белье, здоровье, зелье, кушанье, простонародье, раздумье*.

Как и в парадигмах *Ва* и *Вb*, здесь тоже наблюдаются колебания в принадлежности к определенной парадигме или конкуренция основ:

око	ока	оке (?)	оком	оку					
					очи	очей	очам	очами	оцах
ухо	уха	ухе (?)	ухом	уху					
					уши	ушей	ушам	ушами	ушах
колени	колена	колени	коленом	колену					
					колени	коленей	коленям	коленями	коленях
дерево	дерева	дерева	деревом	дереву					
					деревья	деревьев	деревьям	деревьями	деревьях
платье	платья		платьем	платью	платьев	платьям	платьями	платьях	
крыло	крыла	крыле	крылом	крылу	крыл	крылам	крылами	крылах	
					крылья	крыльев	крыльям	крыльями	крыльях
перо	пера	пере	пером	перу					
					перья	перьев	перьям	перьями	перьях

Парадигма *Ве* изначально включает суффиксы *-а, -е, -у, -ы*, появляющиеся и в других именных парадигмах, и суффиксы *-ой* и *-ою*, которые в других субстантивных парадигмах не встречаются. Последовательная проверка основ на сочетаемость с суффиксами добавляет к этому набору *-0, -ам, -ами, -ах*. К этой парадигме относится 1566 основ, список 30 самых частых слов включает: *анна, баба, бабушка, бабушка, батюшка, вера, вода, голова, госпожа, дама, девушка, дорога, душа, жена, женица, катерина, книга, комната, лиза, минута, москва, надежда, пора, правда, причина, работа, рука, сестра, сила, служба, старуха*. Сравним долю суффиксов в парадигмах *Ва* и *Ве*¹.

	-0	-а	-ам	-ами	-ах	-е	-ов	-ой	-ою	-ом	-у	-ы
Ва	,406	,187	,013	,027	,017	,056	,042	-	-	,102	,084	,057
Ве	,075	,319	,009	,025	,021	,118	-	,077	,023	-	,159	,182

Только суффиксы *-ам, -ами, -ах* показывают небольшие расхождения. У остальных суффиксов различия очень велики, что свидетельствует о существенной функциональной разнице двух парадигм. Об этом же говорит следующая таблица.

¹ Для парадигмы *Ве* подсчет произведен по 106 самым частым основам с общей частотой 283 тысячи словоупотреблений.

Таблица 4.4

Матрица коэффициентов связи (Q) между суффиксами парадигмы Ве

	-0	-а	-ам	-ами	-ах	-е	-ой	-ою	-у	-ы
-0			90	86	93					54
-а										
-ам				95	82					74
-ами					75					47
-ах										
-е								59	64	60
-ой								97		
-у										62
-ы										

Как и раньше, четко выделяется четырехчленный кластер *-0, -ам, -ами, -ах*. Суффикс *-ы* связан с ним и одновременно с более рыхлым кластером *-е, -ой/ою, -у, -ы*. С точки зрения семантической лингвистики такая картина объясняется двузначностью флексии *-ы* (род. п. ед. ч. и им.-вин. п. мн. ч.). Рост доли *-е* объясняется совмещением функций дат. и предл. падежей.

Образование нулевых форм частично повторяет уже отмеченные тенденции других парадигм — сочетания согласных устраняются появлением *-е-*: *свадеб, судеб, усадеб, тюрем, копен, сосен, сестер*; в двух случаях в роли такого гласного выступает *-о-*: *розог, кукол*.

По-другому обстоит дело в основах с конечным *-к-* (по большей части — универсальным суффиксом, присоединяемым к глагольным и субстантивным основам). Здесь можно выявить четкую закономерность. Если конечному *к-* основы предшествует *-ч-, -ш-, -ж-, -й-* или *-ь-*, в нулевой форме находим *-ек* (при этом *-й-* и *-ь-* исчезают). Обнаружено 107 случаев такого рода, ср. *девочек, дочек, привычек, птичек, свечек, речек, собачек, ручек, ласточек, карточек; матушек, девушек, тетушек, старушек, мальчишек, кошек, рубашек, насмешек, старикашек, чашек, лягушек, кумушек; книжек, бумажек, фуражек, ножек, ложек, служек, застежек, денежек, издержек; хозяек, троек, копеек, шаек, скамеек, попоек, построек, индеек, нагаек; маменок, дяденок, нянек, ступенек, полек, дядек, шпилек, мосек*.

При любом другой согласной, предшествующей конечному *к-* основы, в нулевой форме появляется *-ок* (160 примеров): *улыбок, ошибок, рыбок, трубок, юбок; девок, обстановок, лавок, веревок, остановок, торговых, водок, лодок, соседок, поездов, выходок, сходов, загадок, лошадок; сказок, повозок; бутылок, палок, сиделок, прогулок, куколок, елок, тарелок, русалок; немок, рюмок, самок, ужимок, недоимок; француженок, англичанок, служанок, собачонок, амазонок, цыганок, картинок, крестьянок, дворянок, песенок, тропинок, лошадемок, шапок, тряпок, шляпок, скрипок, щепок; кухарок, ярмарок, фигурок, поговорок; колясок, записок, красок, досок, закусок, масок, вывесок, расписок, плясок; теток, шуток, уток, маток, комнаток, брюнеток, пыток, кибиток, решеток, сироток, взяток*.

Помимо универсального основообразующего *-к-* в парадигме Ве отмечено еще два суффикса. Суффикс *-б-* зафиксирован в 15 основах: *служба, свадьба, дружба, злоба, просьба, усадьба, стрельба, пальба, ходьба, тяжба, гульба, мольба, гоньба, похвальба, резьба*. Второй регулярный исход (на 95 % суффикс) *-иц* (150 основ)¹. Назовем 20 самых частых слов этого вида: *красавица, девица, птица, сестрица, разница, улица, умница, пьяница, племянница, половица, царица, покойница, императрица, помещица, убийца, лестница, курица, родственница, благодетельница, ключница, работница*. Основы с конечной *ц-* (за исключением *овц-*) меняют *-ой* и *-ою* на *-ей* и *-ею*; соответствующий вариант можем обозначить Ве-ц.

Парадигма Vf подобно другим «мягким» парадигмам (Vb и Vd) включает два типа: с исходом на «шипящий» и с исходом на другой согласный. В отличие от парадигмы Ве в обоих типах *-ы* меняется на *-и, -ой* и *-ою* на *-ей* и *-ею*.

«Шипящий» подтип включает 100 основ, самые частые из них: *встреча, задача, туча, куча, неудача, притча; юноша, крыша, каша, чаша, польша; сережа, рожка, кожа, невежа, вельможа, сажка, лужка; пицца, роцца, тецца, чацца*. Очень часто *-ш-* выступает как основообразующий суффикс, присоединяемый к основам парадигм Va и Vb: *генеральш-, адмиральш-, губернаторш-, майорш-, мамаш-, папаш-, почтмейстерш-, докторш-, предводителш-, аптекарш-*. Но максимум частоты мы обнаруживаем у имен собственных — вариантов полных личных имен²: *алеша, андрюша, ванюша, глаша, гриша, груша, даша, дуняша, илюша, любаша, маша, милуша, миша, наташа, параша, паша, петруша, саша*.

¹ Лишь в двух основах конечному *ц-* предшествует согласный. Это слова *овца* и *дверца*. В основе *убийц-* в этой позиции оказывается *-й-*.

² Конечно, такая интерпретация возможна только с позиций семантической лингвистики.

Более 100 основ содержат в исходе другие согласные. В этом подтипе в суффиксах последовательно происходит замена *-а* на *-я*, *-у* на *-ю*. Набор суффиксов выглядит так:

-я -ям -ями -ях -е -ей -ею -ю -я

Более 60 основ ограничиваются этим набором: *сударыня, дуня, настя, соня, митя, катя, петя, надя, параня, володя, ваня, федя, варя, тая; беготня, болтовня, возня, дворя, заря, суетня, торговля.*

Что касается нулевой формы, то там появляется либо *-ь* (*княгинь, графинь, барынь, недель, нянь, бурь, пуль, потерь, бань, святынь, героинь, монахинь, минь, кастрюль, заутрень, пилюль, дынь*), либо перед конечной согласной (за которой часто следует *-ь*) вставляется гласный: *песен, обеден, сотен, колоколен, сплетен, башен, часовен, басен, конюшен; барышень, деревень; кухонь; земель, капель, туфель, сабель, петель, кровель, оглобель.*

Парадигма *Vg* объединяет основы с исходом на гласную или на *ь*-. Полный набор суффиксов (*-е, -ей, -ею, -и, -й, -ю, -я, -ям, -ями, -ях*) представлен в слове *идея*, ср. также слова *аглая, свая, стая; аллея, шея, пелагея, палагея, таусея* [sic!], *вируinea, затея, батарея, фея, ворожея, илея, оранжерея, кисея, колея, корея, ливрея, рея, траншея, эпоея; зоя; струя, перепеуя, статуя, сбруя, генуя, чешуя*. В основах с конечным *ь*- обычно нет нулевых форм (*-й*): *марья, дарья, илья, софья, настасья, авдотья, аксинья, анисья, агафья, маланья, прасковья, наталья, гостья, попадьа*. У пяти основ (*семь-, судь-, стась-, свинь-, скаль-*) в нулевой форме *-ь* заменяется на *-ей*, у четырех основ (*игумень-, мантиль-, кель-, колдунь-*) *ь*- заменяется на *-ий*. Громадное большинство основ данной парадигмы имеют исход на *-и*. За исключением 10—12 христианских личных имен (*евпраксия, захария, мария, пульхерия, анатасия, аполлинария, евдокия, евлалия, евламтия, ксения, софия*) на 99 % это общеевропейский материал (в нашем поле наблюдения 180 слов). В порядке уменьшения частоты сюда попадают: *история, юлия, армия, англйя, ассигнация, лидия, россия, полиция, фамилия, академия, партия, компания, фантазия*; и далее по алфавиту: *акация, акция, аллегория, амалия, амбиция, апатия, аристократия, ария, артиллерия, ассоциация, гвардия, губерния, декорация, италия, комедия, комиссия, линия, молния, оказия, поэзия, революция, религия, репутация, станция, теория, трагедия, физиономия, философия, франция, цивилизация, эмилия, энергия* и многие другие.

Парадигма *Vh* обладает следующим основным набором суффиксов: *-ь, -ью, -ию, -и, -ей, -ям, -ями, -ях*.

Доля отдельных суффиксов составляет¹:

-ь	-ью	-ию	-и	-ей	-ям	-ями	-ях	-ьми
,381	,082	,007	,412	,051	,037	,011	,015	,003

Коэффициенты связи покажут сходство *-ью* и *-ию* ($Q = 80$) и существование кластера трех суффиксов

	-ей	-ями	-ях
-ей		86	78
-ями			62
-ях			

Более половины всех основ имеют суффикс *-ост-*, присоединяемый к основам парадигмы *Ca*: *радость, милость, глупость, обязанность, молодость, деятельность, способность, должность, гордость, слабость, наружность, личность, необходимость, возможность, опасность, новость, важность, потребность, бедность, справедливость, крепость, ревность, подлость, малость* и т. д. Восемь слов семантическая лингвистика отнесла бы к числительным (*двадцать, двенадцать, девять, десять, пятнадцать, пять, тридцать, шесть*). Остальные основы кончаются на самые разные согласные: *скорбь, глубь, дробь, прорубь, рябь; любовь, кровь, бровь, свекровь, морковь, ветвь; лошадь, грудь, площадь, тетрадь, исповедь, проповедь, гладь, прядь, заповедь, стерлядь, медь, рухлядь, снеть; грязь, связь, мазь, перевязь; мысль, цель, роль, боль, обитель, пыль, соль, мебель, шинель, дуэль, артель; озимь; болезнь, тень, жизнь, осень, дрянь, казань, зелень, брань, ладонь, казнь, пристань, боязнь, дань, вонь; степь, цепь, поступь; дверь, мать, дочь, сибирь, тварь, утварь, лазурь; надпись, русь, рукопись, смесь, подпись, живопись; часть, честь, страсть, весть, власть, совесть, прелесть, повесть, ненависть, грусть, пропасть; смерть, кровать, память, четверть, печать*. В основах с исходом на шипящие в суффиксах *-я* меняется на *-а*: *дичь, печь, сволочь, желчь, горечь, картечь; роскошь, мышь, фальшь, пустошь, ветошь, плешь; молодежь, дрожь, упряжь; вець, помощь*.

Уникальна парадигма *Vi*, где первообразная основа появляется перед *-л, -ам, -ами, -ах, -е*. Этот кластер суффиксов очень похож на кластеры в парадигмах *Vb, Vc* и *Ve*, иногда осложненные особым основообразующим суффиксом (ср. приведенные выше примеры *друзьям, зятьям, листьям, сыновьям; князьям, мужьям; деревьям, крыльям,*

¹ Подсчет по 78 самым частым основам с общей частотой 140 тысяч словоупотреблений.

СОДЕРЖАНИЕ

Часть 4. Дистрибутивно-статистический анализ в минимальном интервале	5
4.1. Работа ДСА в минимальном интервале на материале корпуса русской прозы	7
4.1.0. Систематизация морфологических данных	7
4.1.0.1. Адъективные парадигмы	8
4.1.0.2. Субстантивные парадигмы	15
4.1.0.3. Глагольные парадигмы	24
4.1.1. Минимальный интервал в русском корпусе	29
4.1.1.1. Частые слова в левой позиции и суффиксы правых слов	32
4.1.1.2. Частые слова в правой позиции и суффиксы левых слов	39
4.1.1.3. Поиск конструкций в зоне слов средней частоты	45
4.1.1.4. Поиск аномалий в конструкциях	50
4.2. Минимальный интервал в английском корпусе	53
4.2.1. Дистрибутивные классы частых слов	54
4.2.2. Комбинаторика дистрибутивных классов	62
4.2.3. Морфология и дистрибутивные классы	69
Часть 5. Дифференциация лексики русской прозы 1850—1870-х годов	75
5.1. Меры лексического сходства частотных словарей	77
5.2. Распределение лексики по основным жанрам прозы 1850—1870-х гг.	80
5.3. Распределение лексики по микрожанрам прозы 1850—1870-х гг.	108
5.4. Распределение лексики по текстам восьми авторов	136
5.5. Расширение круга сравниваемых авторов	168
5.6. Лексические маркеры авторов	171
5.7. Интерпретация авторских лексических маркеров	206
5.8. Авторские маркеры и дистрибутивно-статистическая морфология	216
5.9. Межавторские изоглоссы	231
5.10. Проблема взаимодействия факторов	236
Таблица 5-А. Распределение лексики по основным жанрам	242
Таблица 5-В. Распределение лексики по микрожанрам	327
Таблица 5-С. Распределение лексики по текстам важнейших авторов	441
Таблица 5-Д. Распределение лексики по текстам восьми авторов, родившихся в 1812—1829 гг.	625
Таблица 5-Е. Распределение лексики по текстам девяти авторов, родившихся в 1829—1840 гг.	715
Таблица 5-Ф. Основные параметры распределения лексики по 25 авторам	825
Таблица 5-Г. Ранговый словарь 1500 самых частых лемм	831
Перечень больших таблиц в электронной версии тома 2	839
Литература	843

Научное издание

*Анатолий Янович Шайкевич,
Владислав Митрофанович Андрющенко,
Наталья Александровна Ребецкая*

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

Том 2

Оригинал-макет подготовлен Е. Морозовой
Художественное оформление переплета С. Жигалкина

Подписано в печать 10.10.2016. Формат 60х90/8.
Бумага офсетная № 1, печать офсетная. Гарнитура Times.
Усл. печ. л. 106. Тираж 600. Заказ №

Издательский Дом ЯСК
№ госрегистрации 1147746155325
Phone: **8 (495) 625-35-92** E-mail: Lrc.phouse@gmail.com
Site: <http://www.lrc-press.ru>, <http://www.lrc-lib.ru>

Оптовая и розничная реализация — магазин «Гнозис».
Тел./факс: **8 (499) 255-77-57** E-mail: gnosis@pochta.ru
Костюшин Павел Юрьевич (с 10 до 18 ч.).
Адрес: Москва, Турчанинов пер., д. 4