

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

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

СТАТИСТИЧЕСКИЙ СЛОВАРЬ
ЯЗЫКА РУССКОЙ ГАЗЕТЫ
(1990-е годы)

Том 1



ЯЗЫКИ СЛАВЯНСКИХ КУЛЬТУР
Москва 2008

Издание осуществлено при финансовой поддержке
Российского фонда фундаментальных исследований
(РФФИ)
проект № 08-06-07023



Рецензенты:

доктор филол. наук *Д. О. Добровольский*,
доктор филол. наук *С. Е. Никитина*

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

Ш 12 Статистический словарь языка русской газеты (1990-е годы) / Рос. акад. наук. Ин-т русского языка им. В. В. Виноградова. Т. 1. — М.: Языки славянских культур, 2008. — 592 с., разд. паг. (VIII, 578 с.). — (Studia philologica).

ISSN 1726-135X
ISBN 978-5-9551-0279-5

Настоящий словарь представляет собой композицию трех частотных словарей, за каждым из которых стоит свой корпус текстов: 1) тексты девяти московских и петербургских газет за 1997 г., 2) комплект «Независимой газеты» за 1996—2000 гг., 3) корпус газетных заголовков 1991—2000 гг. Общий объем трех корпусов составляет 50 млн слов текста. В печатной версии словаря представлено 52 тыс. разных слов, в электронной версии словник превышает 140 тыс. разных слов, там же читатель найдет и соответствующий обратный словарь.

Во втором томе словаря будут даны таблицы распределения лексики по газетам, жанрам и темам; электронная версия включит таблицы бинарных словосочетаний.

ББК 83

ISBN 978-5-9551-0279-5

© Авторы, 2008
© Языки славянских культур, 2008

Введение

Жанр частотных словарей когда-то рассматривался лингвистами как сугубо прикладное направление. Один из грандиознейших словарей подобного рода, созданных в докомпьютерную эру, вышел в свет под характерным заголовком — «Словарь учителя...» (Thorndike E. L., Lorge I. The Teacher's Word Book of 30,000 words. N. Y., 1944). Первый частотный словарь русского языка (Josselson H. The Russian Word Count. Detroit, 1953) был адресован преподавателям, он опирался на корпус в 1 млн словоупотреблений (1830–1950) и содержал более 5 тысяч разных лексем. На основе корпуса в 400 тыс. словоупотреблений (литература для детей и юношества) построен «Частотный словарь современного русского языка» Э. А. Штейнфельдт (Таллин, 1963), в нем было представлено 2500 наиболее употребительных слов.

Первый опыт использования компьютера для этих целей материализовался в виде «Частотного словаря русского языка» под ред. Л. Н. Засориной (М., 1977). Словарь построен на базе текстов 1900–1960-х годов общим объемом 1 млн словоупотреблений. Опубликованный словарь включал весь лексический материал (кроме имен собственных). В 1993 г. в Швеции опубликован «Частотный словарь современного русского языка» Л. Леннгрена (Uppsala, 1993). Исходный корпус (1 млн словоупотреблений) содержал в равной доле художественную литературу 1960–1980-х годов и журнально-газетные тексты 1985–1988 годов. К сожалению, в публикацию не включены лексические единицы с частотой 1–9, т. о. в печатном словаре находим около 9 тыс. лексем.

Появление персонального компьютера, сканирующих устройств, развитие Интернета было технологической революцией и, казалось бы, обещало появление все новых и новых частотных словарей. В действительности, прогресс в этой области был не столь быстрым. При росте текстовых корпусов на два (или даже три порядка) автоматизация разметки текста и лемматизации повысили эффективность труда на порядок.

Первым примером такого типа может служить Частотный словарь чешского языка — *Frekvenční slovník češtiny (pod vedením F. Čermáka a M. Křena)*. Praha, 2004. Национальный корпус чешского языка, на базе которого создан этот словарь, включает около 100 млн. словоупотреблений. В словаре представлено более 50 тыс. лемм, начиная с частоты 13.

Одновременно меняются конечные цели, ради которых создаются статистические словари языка. Во главу угла ставится уже не прикладная задача отбора лексики для студентов-иностранцев, но фундаментальная задача описания языка во всем разнообразии жанров, стилей и периодов развития. Первым замечательным образцом такого рода был «*Dictionnaire des fréquences* (P.: Didier, 1971). Соответствующий корпус в 70 млн словоупотреблений включал тексты художественной литературы от 1789 до 1964 г. В итоговых таблицах отражены 4 хронологических среза.

Настоящий словарь должен положить начало серии статистических словарей для разных периодов и разных жанров русского языка. Замысел возник еще в начале 1990-х гг. В те «тошние» для науки годы в Отделе машинного фонда Института русского языка им. В. В. Виноградова РАН удалось вручную ввести в компьютер заголовки нескольких газет 1991–93 гг. («Вечерняя Москва», «Известия», «Комсомольская правда», «Московская правда», «Независимая газета», «Правда», «Сегодня»). Позднее к ним были добавлены заголовки электронных корпусов, упоминаемых ниже. Общий объем корпуса заголовков составляет 1,5 млн словоупотреблений. В первом томе нашего словаря соответствующие частоты даны в правом столбце.

К началу 1997 г. ситуация начала меняться. Появилась возможность подписки на электронные издания газет, что, впрочем, потребовало продолжительной неустанной борьбы с опечатками. Сама эта работа стала возможной благодаря финансовой поддержке РФФИ (гранты 97-06-80100, 00-06-80230, 04-06-80094). Так возник корпус девяти газет 1997 года (в основном, второй половины года): «Известия» (2146 тыс. словоупотреблений), «Литературная газета» (1043), «Московский комсомолец» (1887), «Независимая газета» (3536), «Новая газета» (419), «Правда-5» (737), «Российские вести» (706), «Санкт-петербургские ведомости» (1614), «Сегодня» (2423). В этот корпус были добавлены тексты из националистических изданий 1991–93 гг. («Завтра» и т. п. — 231), из «Московских новостей» 1995 г. (201) и «Литературной газеты» 1994 г. (45). общий объем результирующего корпуса составил 14987 тыс. или круглым счетом 15 млн словоупотреблений. В настоящем словаре этот корпус условно именуется «Корпусом 1997», соответствующие частоты приводятся в левом столбце.

II

Компакт-диски с электронными комплектами «Независимой газеты» стали появляться с 1998 г. Стало возможным появление третьего газетного корпуса — текстов «Независимой газеты» за 1996–2000 гг. общим объемом 35 млн словоупотреблений. Соответствующие данные образуют средний столбец наших таблиц. На материале этого последнего корпуса можно впервые начать анализ краткосрочной лексической динамики, что методически важно для составителей частотных словарей, кроме того, материал этот может заинтересовать историков и политологов.

Таким образом, публикуемый словарь представляет собой композицию трех частично пересекающихся частотных словарей. В левый столбец включено 10% материала среднего столбца (вторая половина 1997 г.), большая часть материала правого столбца присутствует в двух других столбцах. Суммирование данных трех столбцов, следовательно, неправомерно.

Словник

В нашем материале представлено 811 тысяч разных графических слов. Чтобы сделать осмысленным это богатство форм, необходимо свести его к меньшему числу вокабул. В громадном большинстве случаев (но не во всех) вокабула совпадает с леммой, т. е. со словарной формой представления слова. О принципах лемматизации речь пойдет ниже, пока же остановимся на проблеме ограничения объема словника. Для печатной версии словаря вводим следующие правила:

1. Слова русского языка (за исключением имен собственных), их аффиксальные производные и сложные слова со слитным написанием включаются в словарь, начиная с частоты 13.

2. а) Энклитические частицы с дефисным написанием (*-де, -ка, -кась, -от, -с, -таки, -тко, -то*) отделяются от предшествующего слова и включаются в словарь без частотных ограничений. Также разделяются слова с конечным *-другой*.

б) Разделение слова с дефисом не проводится, если частица *-то* присоединяется к местоименному слову с начальным *к-* (*з-*) или *ч-*.

в) Сложные числительные с дефисным написанием разделяются на два слова. Также разделяются дефисные написания с названиями месяцев.

г) Не разделяются на части графические слова *вообще-то, все-таки, наконец-то, ну-ка, ну-с, ну-тка, опять-таки, потому-то, столько-то, сякой-то, так-то, такой-то, то-то, туда-то, тут-то*.

3. Имена собственные включаются в словарь, если их частота превышает 29. Этот же порог действителен для аббревиатур организаций, компаний и фирм, всевозможных мероприятий с цифровым компонентом (*Ту-130, Формула-1* и т. п.).

4. Леммы, записанные латиницей, включаются в словарь, начиная от частоты 5.

5. В словарь включены и числа, если их частота превышает 99.

Лемматизация

Строгое разделение частей речи потребовало бы изощренной программной системы или ручного обследования миллионов контекстов. Особенно трудоемко разделение прилагательных и причастий, в несколько меньшей мере прилагательных и существительных. Далеко не все эти задачи имеют однозначное решение.

При составлении данного словаря мы следовали гибкой политике, иногда жертвуя грамматической чистотой в пользу лексической семантики. Синтетические компаративы и суперлативы выделялись в особые вокабулы, равно соотношенные с прилагательными и наречиями. Деадъективные формы на *-о* считались особыми вокабулами (в них может быть скрыт какой-то процент кратких форм среднего рода прилагательных).

Стремление к посильной семантической дифференциации заставляло иногда расщеплять традиционные словарные статьи обычных словарей. Так, разделены формы числа существительных *Бог* и *боги, выбор* и *выборы, круг* и *круги, многое* и *многие, новости* и *новость, рамка* и *рамки, сведение* и *сведения, цвет, цветок* и *цветы*. Трактуются как особая вокабула форма *образом*. Отдельными вокабулами стали *болит* (*болело*), *давай(те)*, *кажется, придется* (*пришлось*), *приходится* (*приходилось*), *разумеется, следует* (*следовало*), *сообщается, спрашивается, стоит*

(*стоило* — как модальное слово), *хватает* (*хватало*) и *хватит* (*хватило*). При таком расщеплении учитывались как семантические, так и частотные соображения. Большинство выделенных глагольных вокабул встречаются намного чаще суммы всех остальных форм соответствующего глагола.

Особыми вокабулами являются многие краткие формы прилагательных: *велик, намерен, нужен, обязан, принято, равен, согласен, таков, убежден*. Довольно часто в отдельные вокабулы выделялись потенциальные причастия (частично адъективированные или субстантивированные): *действующий, заведующий, интригующий, командующий, кричащий, курящий, любящий, манящий, мертвящий, моющий, мыслящий, мятущийся, обжигающий, облитерирующий, ободряющий, объемлющий, окружающий, освежающий, отдыхающий, отравляющий, отягчающий, ошеломляющий, палящий, подавляющий, подобающий, подрастающий, подходящий, порочащий, последующий, потрясающий, правительствующий, правящий, практикующий, предержащий, предстоящий, предшествующий, предыдущий, преходящий, приходящий, прилегающий, прилежащий, приличествующий, проезжающий, проникающий, присутствующие, профилирующий, процветающий, пьянящий, решающий, руководящий, соответствующий, страждущий, странствующий, текущий, трудящийся, тяжущийся, угрожающий, удручающий, ужасающий, укоряющий, упреждающий, успевающий, устрашающий, учащийся, фашиствующий, хрустящий, хулиганствующий, цветущий, чарующий; минувший, прошедший; обвиняемый, предполагаемый; направленный, обеспеченный, обреченный, обусловленный, основанный, построенный, предназначенный, представленный, предусмотренный, расположенный, распространенный, связанный, сокращенный, указанный.*

В какой-то мере разделение на вокабулы проводилось в рамках одной части речи исключительно по семантическим соображениям. Таковы *лев, лев* (болг.) и *Лев; среда и среда* (день недели), *статья и статья* (журн.).

Словосочетания

Отличительной чертой настоящего словаря можно считать включение в него в качестве вокабул не только отдельных слов, но и словосочетаний. Нам известен только один частотный словарь, дающий сочетания слов (*collocations*). Это Allen S. e. a. *Frequency Dictionary of Present-day Swedish based on newspaper material*. Uppsala, 1970–75. Статистике словосочетаний посвящен 3-й том словаря, где приводятся все бинарные словосочетания с частотой 2 и более (включая самые тривиальные грамматические сочетания). Словарь Аллена построен на корпусе 1 млн словоупотреблений.

В таблицу 1 настоящего словаря включено несколько десятков словосочетаний, таких, как: *в качестве, в конечном счете, в конце концов, в связи с, в противном случае, в случае, в соответствии с, в состоянии, в сущности, в течение, в том числе, в угоду, в ходе, в частности, вверх тормашками, во избежание, вряд ли, до свидания, до сих пор, друг друга, за неимением, за счет, и так далее, и тому подобное, исполняющий обязанности, к сожалению, круглый стол, между тем, может быть, на плаву, на скаку, на протяжении, населенный пункт, настоящее время, по крайней мере, по поводу, по сравнению с, по существу, повестка дня, потому что, права человека, с поличным, сломя голову, стало быть, судя по, Счетная палата, табель о рангах, так как, так называемый, так что, таким образом, тем не менее, тем самым, то есть, только что, тот самый, точка зрения, Филькина грамота, что касается.*

Во втором томе настоящего словаря будут представлены все бинарные словосочетания с частотой 3 и более. Приведем несколько (лемматизированных) примеров с левым прилагательным. Числа указывают на частоту словосочетания, числа в скобках - частота прилагательного.

кричащий [82] заголовок 4, проблема 3, противоречие 3
моющий [21] средства 11
мятущийся [46] душа 9
мыслящий [292] люди 27, политик 9, человек 9, лидер 5, часть 5 интеллигенция 3
отравляющий [88] вещество 50, газ 5
ошеломляющий [87] успех 15 впечатление 8 победа 3

Обратимся к более частым словам:

IV

действующий [5766]:

ныне 30, сейчас 15, сегодня 11, уже 5;
законодательство 457, конституция 380, закон 178, соглашение 18, устав 11, документ 10,
договор 9, УК 9, программа 8, контракт 5, редакция 5;
система 54, модель 22, механизм 21, норма 20, порядок 18, режим 11, ставка 11, правило 10,
схема 8, бюджет 8, санкции 4, тариф 4;
президент 400, губернатор 231, председатель 174, глава 107, мэр 35, премьер 18,
руководитель 13, премьер-министр 9, лидер 9, министр 9, градоначальник 5;
власть 96, политик 49, правительство 37, депутат 30, парламент 22, кабинет 14,
руководство 7, состав 6;
орган 35, структура 13, круглый стол 5;
сотрудник 15, член 12, генерал 7, офицер 6, чиновник 5;
чемпион 10;
храм 11, церковь 4;
предприятие 18, газопровод 8, шахта 7, АЭС 6, банк 5, месторождение 5, выставка 5,
объект 4, станция 4, блок 4, скважина 4;
субъект 7;
фактор 7;
армия 34;
лицо 415.

Приведенные примеры сгруппированы по смыслу и в своей совокупности семантически характеризуют данное прилагательное. Обратим внимание на сравнительно новое словоупотребление — *действующий чемпион*. Мы видим здесь также кандидата во фразеологизмы - *действующая армия* и безусловный фразеологизм - *действующее лицо*.

Прилагательное *действующий* лишено каких-либо эмоциональных коннотаций. Напротив, прилагательное (или причастие) *правящий* ярко окрашено стилистически, и правые компоненты словосочетаний это демонстрируют.

правящий [2504]:

партия 491, коалиция 336, элита 255, режим 123, круги 119, класс 113, верхушка 60,
альянс 41, слой 39, большинство 38, группа 20, группировка 18, блок 18, династия 15,
бургомистр 12, олигархия 11, архиерей 10, кабинет 10, лагерь 10, ныне 10, номенклатура 9,
семья 8, команда 7, бюрократия 3, епископ 6, курс 6, структура 6, власть 5, президент 5, клан 4,
монарх 4, истеблишмент 3, пока 3, триумвират 3.

Исследователь-лингвист многое почерпнет из статистики, казалось бы, самых простых сочетаний. Вот два примера предложного управления:

на Украине	2504	в Украине	142
на Украину	618	в Украину	25
на Охотном	213	в Охотном	48

Мы видим, что «Независимая газета» в 1990-х годах держится традиции предложного управления словом *Украина*, но явно отказывается от традиционного московского *в Охотном ряду*.

Хронология

В словарь включена хронологическая таблица текстов «Независимой газеты». Ниже приводится несколько примеров различных тенденций динамики частоты:

	1996	1997	1998	1999	2000
Монотонное нарастание частоты					
Евросоюз/ЕС	504	676	997	974	1663
компания	2904	3656	4225	4117	4759
холдинг	37	77	229	270	509
СМИ	776	1155	1289	1506	2041
Кремль	1164	1338	1341	1636	2427
чиновник	938	1225	1298	1069	1308
номинация	194	291	314	331	341
суд	2813	2880	3384	3827	4389
(ген)прокуратура	907	986	1190	1370	1957
экстремист/-зм	283	224	480	749	889
фигурант	18	18	36	69	53
Березовский	173	458	725	790	854
Монотонное падение частоты					
СССР	2976	2160	1821	1712	1874
оппозиция	2736	2440	2108	1723	1668
революция	1052	927	841	763	772
джаз	160	144	43	79	68
Пик частоты в 1996 г.					
Ельцин	8485	5678	5302	4238	2069
консилиум	51	14	9	17	6
теннис	93	54	58	57	40
Гайдар	573	467	518	313	148
КС	499	497	339	147	192
Лебедь	3360	913	1105	414	212
Зюганов	3165	698	1032	837	763
Пик частоты в 1997 г.					
секвестр	2	323	53	10	14
Диана	33	164	58	40	38
празднование	152	273	211	200	245
Пик частоты в 1998 г.					
доллар	3692	3916	5290	4443	4752
рубль	3327	3943	5019	3377	3337
правительство	9458	10810	13916	10652	9455
отставка	1172	1356	1846	1750	1222
Резкий подъем частоты в 1998 г.					
олигарх	3	27	496	238	842
Пик частоты в 1999 г.					
МВФ	651	557	1098	1729	781
дефолт			90	230	136
Примаков	788	1115	2497	3637	614
Степашин	79	86	439	2299	297
импичмент	82	94	454	758	70
одномандатный	43	40	108	391	109
Косово	8	8	1164	2812	1068
Пушкин	723	722	773	1488	541

VI

Пик частоты в 2000 г.					
Путин	11	14	189	2336	10133
Касьянов	11	7	23	203	1260
Гусинский	49	66	96	118	1151
Кадыров	6	3	36	46	325
политтехнолог	1			12	63
святой	643	581	531	560	874
гимн	114	128	118	113	435
вертикаль	128	102	115	101	424
дзюдо	12	3	4	3	46
погибнуть	1146	964	1135	1242	1937
подводный	179	176	256	171	815
Пики частоты в четные годы					
олимпиада	240	50	155	86	153
Снижение частоты в 1997-98 гг.					
президент	18469	16641	17313	14600	18239
выборы	8553	3855	6117	8485	7263
соперник	484	222	311	320	418
патриарх	423	342	273	337	437
Чечня	5077	4314	2349	3920	4418
боевик	1442	425	335	1397	2068
солдат	711	581	551	759	735
бандит	234	277	196	440	425
Изменения частот в связи с внутригазетными структурными и стилистическими обстоятельствами					
балет	449	680	650	761	1273
наука	2350	1936	1578	1321	1264
кино	1100	1146	661	705	841
книга	3300	1851	1886	1592	1700
архитектура	420	507	158	125	189
г-н	3805	4040	3171	1432	982
вы	7701	6741	6560	5107	4927
мой	4503	4046	3554	2828	2735
бы	15055	13341	13345	12092	12167
ни	9313	7738	8367	7938	7846
странно/ый	1108	970	835	828	761
корова	218	80	124	86	84
ну	1348	1127	919	725	687
оньй	87	63	36	54	32
милый	272	209	175	117	88
совесть	351	356	258	193	205
Освоение заимствований					
перфоманс	63	32	29	9	1
перформанс	5	9	18	20	17
рг	4	5	30	17	33
пиар			1	4	74
пиаровский			5	17	50

Возможности внешнего и внутреннего сравнения

В настоящее время доступен только один источник для внешнего сравнения нашего корпуса: это электронный частотный словарь, составленный С. А. Шаровым на базе строящегося Национального корпуса русского языка. Корпус быстро растет, растет и частотный словарь. В нашем распоряжении есть версия начала 2006 г. с базой в 16 млн. словоупотреблений, которую можно сравнить с нашим левым столбцом. (Корпус Шарова включает преимущественно тексты художественной литературы, в том числе и старой). Вот относительные частоты некоторых слов в обоих словарях (на миллион словоупотреблений):

	Шаров 1997			Шаров 1997	
а	9808	6204	в	29028	41058
весь	8605	3243	год	2042	4058
и	36359	30908	для	2250	3634
к	5740	5002	который	3289	5264
не	20690	14502	Москва	421	1604
но	5673	4333	новый	840	1933
у	5033	2926	о	3518	4896
я	15631	4001	первый	1293	1966
баба	138	13	по	5735	7796
бабушка	120	35	политический	141	702
барахло	11	1	правительство	123	1225
бежать	222	60	президент	144	1737
беззвучно	16	1	российский	150	2074
белый	493	220	Россия	359	3033
берег	240	109	страна	467	1716
береза	24	4	активный	45	115
беспокоиться	45	13	аналогичный	18	80
бесшумно	24	1	аргумент	20	60
бить	146	51	арестовать	38	89
блестеть	46	3	аспект	13	33
Бог	377	149	банк	138	674
болтать	37	4	безопасность	56	380
боль	192	59	безусловно	29	76
большой	1631	855	Белоруссия	9	187
бормотать	46	3	беспредел	3	14
босой	24	3	бизнес	43	164
ботинок	54	11	биржа	7	53
бояться	338	112	благодаря	55	137
брат	272	119	ближайший	81	277
бровь	72	6	более	618	1299
бродить, брести	71	17	большинство	112	310
бросить, бросать	344	114	бомж	10	26
брюки	57	13	борьба	134	383
бумажка	53	15	будущий, -ее	202	411
бутылка	170	59	бывший	174	492
бык	80	10	бюджет	33	551
валяться	52	8	бюджетник	1	29
варить	64	5	валюта	21	94
ваш	668	343	валютный	11	89
вдоль	100	26	ведомство	17	167
вдруг	713	154	версия	40	140
ведро	56	10	верховный	26	109
ведь	884	533	весьма	120	285

VIII

ведьма	33	5	ветеран	18	77
вежливо	37	7	вещание	3	27
везти	61	30	взаимоотношение	9	69
веревка	51	9	визит	31	246
веселый, -о	209	51	вице-премьер	-	267
ветер	209	60	вклад	13	115
вечер	391	258	включая	30	87
взять, братъ	1003	515	владелец	42	111
видеть	1062	362	власть	279	1047
вниз	158	23	военный	239	606
вода	618	266	возвращение	44	106
воздух	276	103	возглавить, -лять	25	176
волноваться	70	18	вооруженный	41	193
вон	146	21	вопрос	511	1229
вот	2205	859	воскресенье	37	89

Общее впечатление от сравнения двух списков: В корпусе Шарова намного больше глаголов, местоимений 1 и 2 лица, слов, обозначающих действие, имеющих отношение к психологии, зрению, слуху, природе. Множество других маркеров корпуса подтверждают это общее впечатление. Не только *бежать*, но и *броситься, быстро, быстрый, вернуться, возвращаться*. Не только *бык*, но также *амбар, баран*. Наряду со словом *белый* видим также *алый, багровый, бурый*. На советское прошлое (на войну, в частности) указывают *батальон, бинокль, боец, бригадир, взвод, авоська*. Рядом со словом *Бог* находим *ад, ангел, вера, верить*. Некоторые слова корпуса пахнут XIX веком, ср. *ах, барин, бледный, борода, братец, взгляд, взглянуть, вздохнуть, возражать/-зить, воскликнуть, возле*.

В газетном корпусе 1997 г. преобладают имена, конкретные лексемы легко объясняются современной тематикой российских газет. Такой эффект объясняет слишком большим стилистическим и хронологическим расстоянием между двумя корпусами. Примером внутреннего сравнения газет могут служить стилистически резко различающиеся «Московский комсомолец» и «Независимая газета» (относительная частота на 1 млн словоупотреблений).

	МК	НГ		МК	НГ
аж	116	12	амбиция	19	43
алкоголик	24	3	бандформирование	3	20
апартаменты	20	3	большевизм	2	8
баба	32	8	брифинг	14	25
бандит	123	49	буржуазия	2	7
бардак	8	2	глобализация	1	19
блокбастер	4	1	глобальный	37	71
бойфренд	8	1	декларация	44	74
болельщик	41	11			
бомж	42	10			
босс	41	6			
бутик	14	1			
вдруг	250	100			
вырезвитель	15	1			
граффити	11	1			
девушка	292	51			

Второй том словаря покажет внутреннюю дифференциацию лексического материала по газетам, жанрам и темам. Как уже указывалось выше, в электронной версии будут представлены бинарные словосочетания.

Распределение лексем по трем корпусам текстов

А — корпус 1997 года — 15 млн слов текста;
 В — корпус «Независимая газета» 1996–2000 гг. — 35 млн слов текста;
 С — корпус заголовков 1991–2000 гг. — 1,4 млн слов текста

А

а	91783	4358		Абдуллаев	20	70	2
	178444			Абдулло	95	386	5
а (междометие)	43	37	1	Абдулов	72	91	5
А.	2825	134		Абдумалик	17	97	4
	6624			Абдуррахман	1	36	
а-3	10	49	5	Абель	18	46	7
а-ля	63	138	6	Абеляр	43	48	1
а-студио	17	32		аберрация	10	38	1
АА	8	33		абзац	88	320	4
Аарон	10	31	1	Абиев	3	40	
аб	35	30	1	Абиола	4	32	2
Абаев	3	51	2	абитуриент	120	125	18
абажур	16	18	1	Абишевич	5	35	
абазин	3	54	3	абонемент	115	398	5
абазинский		30	2	абонементный	24	31	1
Абай	5	36	1	абонент	236	122	1
Абакан	22	35	1	абонентный	39	22	
Абакумов	9	38	2	абонентский	33	40	2
Абалкин	24	84	10	абориген	79	134	14
Абача	9	51		аборт	164	350	30
Абашидзе	72	348	10	абракадабра	7	23	2
Аббадо	18	77	10	Абрам	32	197	10
Аббас	23	86	6	Абрамов	84	213	8
Аббасов	14	87	4	Абрамович	159	459	14
аббат	13	30	1	Абрамцево	9	39	4
аббатство	36	9	1	абрамцевский	5	16	1
аббревиатура	69	1	4	абрикос	13	8	
абд	2	13		абрис	9	19	1
Абдалла	16	62	2	Абсандзе	13	48	
Абдель	8	114	3	абсолют	51	93	5
Абдрашитов	54	95	4	абсолютизация	5	32	1
Абдул	44	132	4	абсолютизировать	4	37	
Абдул-вахед		30	2	абсолютизм	13	46	3
Абдулаев	7	32		абсолютистский	5	14	
Абдулазиз	7	55		абсолютно	1617		22
Абдулатипов	239	363	26			3542	
Абдулла	30	236	4	абсолютность	3	16	
Абдулладжанов	19	175	3				

	А	В	С		А	В	С
абсолютный	709		35	аварский	24	81	
		2156		авачинский	13	5	1
абсорбция	1	28		авва	10	17	1
абстрагироваться	15	58		Аввакум	11	38	1
абстрактно	11	43		Аввакумов	54	44	3
абстрактный	142	516	8	авгиев	4	15	1
абстракционизм	6	28	2	авгур	4	33	1
абстракционист	6	19	2	август	4080		240
абстракция	33	139	7			7786	
абсурд	199	537	40	августейший	27	42	5
абсурдистский	8	48		Августин	14	58	3
абсурдно	41	109	2	августовский	164	588	35
абсурдность	48	127	2	Авдеев	23	123	3
абсурдный	145	366	6	Авель	22	36	3
Абу	50	207	5	Авен	15	135	8
Абу-Даби	32	114	13	авеню	27	33	4
Абубакар	1	45		Авербук	33	5	
Абубакаров	2	32		Аверинцев	15	36	1
Абульфаз	17	90	13	Аверченко	8	29	1
абхаз	84	295	21	авиа	19	29	1
абхазец	17	45	2	авиабаза	52	134	4
Абхазия	1129		341	авиабилет	66	52	12
		5125		авиабомба	13	52	5
абхазо-грузинский	6	19	1	авиавыставка	2	14	3
абхазский	497		85	авиадвигатель	21	74	3
		2141		авиадивизия	1	15	
Авалиани	13	31		авиадиспетчер	55	60	41
авангард	268	821	77	авиазавод	42	97	4
авангардизм	5	31	2	авиакатастрофа	107	159	78
авангардист	41	116	6	авиакомпания	507		38
авангардистский	11	47	1			1016	
авангардный	56	257	3	авиакомплекс	9	21	1
аванс	143	207	6	авиаконструктор	21	23	5
авансирование	14	45	1	авиакор	8	70	6
авансировать	12	21		авиакосмический	54	102	7
авансовый	35	55	1	авиалайнер	89	136	10
авансцена	64	169	8	авиализинг		13	
авантюра	82	288	15	авиалиния	143	317	16
авантюризм	25	59	3	авианалет	4	40	4
авантюрист	75	172	16	авианесуший	29	34	3
авантюристический	1	26	1	авианосец	88	163	16
авантюрный	43	102	1	авианосный	12	22	
аварец	50	134		авиаотряд	17	17	1
аварийно-				авиапарк	6	16	
-восстановительный	3	14		авиапассажир	15	18	5
-спасательный	35	67	1	авиаперевозка	41	96	5
аварийность	34	52	1	авиаперевозчик	31	50	3
аварийный	329	494	15	авиаполк		16	1
авария	1110		245	авиапредприятие	54	64	3
		1729		авиапромышленность	56	99	11

	А	В	С		А	В	С
авиарейс	27	67	8	авто-	14	14	
авиаремонтный	29	24	3	автоавария	11	14	9
авиасалон	100	175	42	автобаза	27	17	3
авиасообщение	2	19	2	автобан	14	27	4
авиастар	11	35	2	автобанк	46	50	6
авиастроение	26	73	3	автобензин	7	15	5
авиастроитель	29	78	15	автобиографический	42	116	4
авиастроительный	20	74	1	автобиографичный	6	18	
авиатехника	73	120	7	автобиография	34	100	3
авиатор	67	112	24	автобус	1104		118
авиатранспорт	13	25	3			1212	
авиатранспортный	4	23		автобусный	149	105	10
авиаудар		93	4	Автоваз	210	342	58
авиационно-				автовладелец	97	43	13
-космический	26	64	5	автовокзал	20	22	1
-технический	3	15		автогигант	40	35	8
авиационный	646		35	автогонка	49	59	4
		1366		автогонщик	16	5	2
авиация	668		80	автограф	268	419	32
		1917		Автодор	80	44	13
авиашоу	17	14	7	автодорога	133	228	7
авиаэкспорт	4	14		автодорожник	8	13	1
авизо	39	29	9	автодорожный	34	38	2
авизовка	13	1		автодром	16	34	2
Авилов	40	9	1	автозавод	190	189	19
Авиньон	76	102	17	автозаводец	22	7	4
авиньонский	43	78	4	автозаводской	13	13	1
авионика	22	45	1	автозаправка	33	27	4
Ависма	7	30	1	автозаправочный	46	76	
Авода	16	43		автозапчасти	24	13	1
авось	79	148	19	автоинспектор	15	3	
авоська	23	32	3	автоинспекция	18	16	1
АВПК	48	58	3	автокатастрофа	155	148	46
Авраам	40	143	3	автокефалия		42	
аврал	39	42	13	автокефальный	2	33	2
авральный	35	53	1	автоколонна	20	70	1
Аврора	108	106	23	автокомбинат	24	13	4
австралиец	128	102	17	автократический	4	25	
австралийский	180	234	19	автократия	4	15	
Австралия	429	562	68	автолайн	20	3	
австриец	69	134	9	автолюбитель	128	60	25
австрийский	296	754	57	автомагистраль	54	138	
Австрия	469		91	автомат	829	960	89
		1220		автоматизация	36	51	2
австро-венгерский	8	42	1	автоматизированный	89	120	3
Австро-Венгрия	10	63	2	автоматизм	10	26	
АВТ	108	63		автоматика	40	33	3
Автандил	24	50	1	автоматически	325	863	5
автаркия	3	34	3	автоматический	307	541	8
авто	153	76	18	автоматный	63	83	8

	А	В	С		А	В	С
автоматчик	59	53	5	автостоянка	147	83	12
автомашина	609	582	14	автострада	37	70	5
автомобилестроение	48	101	2	автотехника	11	17	2
автомобилестроитель	23	26	2	автотранспорт	218	323	8
автомобилестроительный	16	22		автотранспортный	60	73	2
автомобилист	180	91	24	автотрасса	76	126	7
автомобиль	2961		249	автофургон	8	16	
		3185		автохозяйство	19	12	
автомобильный	734		55	автохтонный	1	22	
		1130		автоцистерна	28	27	2
автономизация	9	30		автошоу	6	13	2
автономия	411		116	автоэкспедиция	1	15	
		2032		Автурханов	20	47	3
автономно	20	66		автуры	1	25	
автономность	7	50		авуары	15	32	2
автономный	494		46	АГ	24	90	4
		1903		ага (междометие)	44	56	4
автоответчик	26	25		Ага Хан	9	64	
автопарк	52	27	4	Агапов	68	62	1
автоперевозчик	22	29	5	Агасси	54	26	4
автопилот	9	14	1	агат	26	23	2
автопоезд	52	11		Агата	41	55	3
автопортрет	17	124	11	Агафонов	67	36	1
автопредприятие	15	11	1	Агафья	5	51	4
автопробег	22	62	17	Агдам	4	40	8
автор	4601		162	агдамский	3	24	1
		12314		Агджа	16	37	5
авторизация	17	4		Агеев	30	41	1
авторизованный	7	18	2	Агейчев		36	
авторитаризм	20	304	9	агент	889		184
авторитарность	13	39				1261	
авторитарный	127	625	11	агентский	7	15	
авторитет	1096		90	агентство	1832		88
		2658				4128	
авторитетнейший	16	35		агентура	73	133	6
авторитетно	21	38	1	агентурный	67	161	11
авторитетность	2	23	1	агинский	16	23	1
авторитетный	381		10	агитатор	62	107	11
		1119		агитационный	84	255	6
авторник	4	22		агитация	161	857	35
авторский	616		27	агитировать	70	210	28
		1393		агитка	16	57	3
авторство	77	217	8	Агитпроп	26	58	5
авторучка	22	19	2	АГК		51	2
авторынок	17	10	2	Агламишьян		59	
автосалон	58	41	15	агнец	21	38	6
автосборочный	14	10		агон		19	1
автосервис	36	37	6	агонизировать	16	32	1
автослесарь	22	4	3	агония	48	126	17
автоспорт	30	8	1	аграрий	170	581	53

	А	В	С		А	В	С
аграрник	24	111	8	адекватный	259		15
аграрно-					1201		
-демократический	5	51		Аделаида	10	71	2
-промышленный	10	42	1	Адель	10	30	1
аграрный	467		73	Аденауэр	9	53	
		1488		адепт	56	157	1
аграф	8	29		Аджария	99	497	22
агрегат	126	163	1	аджарский	34	144	2
агрегатный	6	15		Адзельо	1	30	1
агрессивнее	4	15	1	адзурра	20	3	
агрессивно	67	141	2	адидас	95	11	27
агрессивность	108	202	4	Адлер	35	57	1
агрессивный	353	901	24	административно	17	30	
агрессия	269		54	административно-			
		1577		-командный	6	67	
агрессор	61	291	11	-территориальный	5	241	4
Агриппина	11	31		-хозяйственный	2	22	
агрокомплекс	13	12	1	административный	1092		54
агроном	30	83	2		3135		
агрооблигация	10	13	1	администратор	154	391	9
агропласт		18		администрация	5356		369
Агропромбанк	112	93	11		13878		
агропромышленный	150	403	5	администрирование	16	89	5
агросфера		16		адмирал	398	746	98
агрофирма	15	46	2	адмиралтеец	13	4	1
ад	225	626	65	адмиралтейский	121	39	3
Ада	5	65	1	адмиралтейство	57	25	6
адажиетто	4	14		адмиральский	16	10	1
адажио	11	52		адов	21	12	1
Адам	133	326	18	Адольф	81	258	14
Адамишин	91	103	08	АДП	10	48	
Адамия	24	66	1	АДЛМ	8	32	
Адамкус	22	124	7	АДР	16	157	4
Адамов	4	246	17	адреналин	17	24	
Адамович	56	125	2	адрес	1944		121
Адамс	112	131	7		4006		
Адамчук	114	40		адресат	135	178	9
адаптационный	10	18		адресно	10	24	
адаптация	151	374	5	адресность	11	29	1
адаптированный	18	96	1	адресный	93	197	
адаптировать	44	107		адресовать	267	538	4
адаптироваться	70	137	5	адресоваться	21	58	2
адат	12	33		Адриан	80	95	1
адвокат	1203		160	Адриатика	13	47	6
		2253		адриатический	12	46	1
адвокатский	68	108	2	адский	107	158	15
адвокатура	34	93	3	адъютант	37	63	3
Аддис-Абеба	4	69	8	адыг	9	112	6
адекватно	107	365	7	адыгеец	2	50	
адекватность	11	91	2	адыгейский	3	42	

	А	В	С		А	В	С
Адыгея	33	287	19	АИФ	15	39	2
адыгский	1	28		ай	34	63	14
адюльтер	5	16	3	ай-би-эм	32	14	4
аж	576	417	4	ай-си-эн	28	3	
Ажгихина	13	104	2	Айвазовский	64	161	13
ажиотаж	225	377	24	Айги	8	49	2
ажиотажный	36	71	6	Айдан	7	38	2
ажурный	26	40	1	Айдан-галерея	6	29	2
аз	21	24	6	Айдид	6	49	12
Азамат	13	55	2	Айзенберг	7	24	3
азарт	151	167	10	Айк	5	29	2
азартно	24	32	1	айкидо	47	5	3
азартный	91	111	13	Айова	22	25	2
азат	20	30		Айрапетова	17	199	
азбука	44	160	7	Айрикан	10	66	2
азбучный	12	37	3	айсберг	82	115	12
Азербайджан	1892		499	Айседора	23	37	5
		8722		Айтматов	39	108	17
азербайджанец	75	534	16	Айюб	15	23	1
азербайджано-иранский	2	13	1	АК	91	319	7
азербайджано-российский	7	24	1	АК&М	3	39	1
азербайджанский	799		101	Ак-барс	35		2
		2831		академгородок	25	27	3
азери	33	94		академизм	7	52	7
азиат	43	81	6	академик	773		186
азиатский	364		78			2008	
		1150		академик-секретарь	3	39	
азиатско-				академический	414		39
-тихоокеанский	126	329	7			1181	
азиатчина	6	20	1	академичность	4	17	
Азиз	72	227	26	академичный	6	20	
Азимов	11	142	1	академия	1749		185
азимут	10	60	4			3930	
Азия	1172		273	Акаев	76	698	48
		4184		Акакий	24	83	3
АЗЛК	61	54	14	акафист	1	13	1
Азнавур	17	48	6	акация	14	25	2
Азов	25	84		АКБ	62	62	2
азовский	46	124	2	Акбар	50	300	5
Азольский	53	35	4	акваланг	23	19	4
азорский	2	21		аквалангист	15	16	2
азот	85	105	7	аквапарк	29	8	3
азотный	12	31	4	акварель	39	199	6
азс	136	139	12	акварельный	8	56	1
АзССР	15	35		аквариум	93	152	21
АЗХХ	6	31		аквариус	20	6	1
азы	33	98	5	акватория	143	303	6
аи	20	9	1	Акежан	28	141	3
Аида	55	191	9	аки	21	38	2
аист	26	18	11	аким	36	78	3

	А	В	С		А	В	С
Акимов	36	81	2	активизация	265		25
акклиматизация	8	16				1203	
аккомпанемент	49	118	6	активизировать	153	721	50
аккомпаниатор	11	24	1	активизироваться	117	420	76
аккомпанировать	30	63		активист	262	881	30
аккор	6	18	3	активистка	32	104	8
аккорд	72	148	13	активнее	145	357	26
аккордеон	11	42	1	активнейший	24	61	
аккредитация	141	198	3	активно	1432		54
аккредитив	14	12	1			4353	
аккредитованный	105	242	2	активность	856		109
аккредитовать	13	13				2731	
аккумуляирование	7	39	2	активный	1724		90
аккумуляированный	13	38				5463	
аккумуляировать	51	138	2	активы	292	823	18
аккумуляироваться	15	42		актовый	37	58	1
аккумуляитор	44	47	4	актриса	758		57
аккумуляиторный	13	32				1623	
аккумуляция	5	18		актуализация	8	40	1
аккурат	93	112	2	актуализировать	8	59	1
аккуратнее	9	16	1	актуализироваться	3	34	
аккуратно	210	263	4	актуальнее	4	33	4
аккуратность	16	47	2	актуальнейший	9	25	
аккуратный	85	118	4	актуально	56	220	4
Акмола	134	111	10	актуальность	105	437	3
акмолинский	3	16		актуальный	484		39
Акопов	17	80				1792	
акр	22	15		актюбинский	7	15	
акриламид	15			акула	172	199	38
акробат	40	47	4	Акунин		48	3
акробатика	17	24	1	акустика	31	68	3
акробатический	19	35	2	акустический	45	90	
Акрополь	14	32	3	акушерка	18	15	1
аксакал	16	36	7	акушерство	16	11	2
Аксаков	37	85	3	акцент	287	984	43
Аксененко	48	643	48	акцентирование	2	14	
Аксенов	82	168	12	акцентировать	55	279	2
аксес		17		акцентироваться	4	29	
аксессуар	64	69	3	акциз	375	652	23
аксиома	77	209	7	акцизный	168	193	3
Аксючиц	21	86	10	акционер	1078		116
акт	1726		125			2063	
		5439		акционерный	811		38
актер	2193		149			1384	
		5415		акционирование	150	223	18
актерски	2	17		акционировать	54	81	11
актерский	400	974	16	акционироваться	25	18	7
актерство	11	25	3	акция	6279		643
актив	142	352	5			13587	
				Акчурин	43	45	

	А	В	С		А	В	С
алагирский	8	17		Алетейя	6	38	
Алан	96	301	10	Алехин	65	32	5
Алания	294	252	13	Алеша	91	104	10
аланы	2	21	1	Алешин	16	44	
албанец	55		78	Алешковский	14	81	2
		1382		Алжир	153	527	102
Албания	191	819	115	алжирец	24	44	4
албанский	90	938	49	алжирский	50	185	10
алвазиец	13			Али	135	392	14
алвазийский	20			алиби	33	41	1
Алвазинск	35			Алиев	570		171
Алвазия	77				2282		
алгебра	17	33	4	Алиевич	19	49	1
алгоритм	24	137		алиевский	3	14	
Алдамов	8	35		Алик	28	158	1
Алданов	7	61	2	алим	2	14	
Алевтина	65	156	1	алименты	42	21	7
Алези	33	1	4	Алимов	31	19	
Алейников	20	41	2	Алина	58	88	5
Алекперов	60	258	6	Алиса	139	187	19
Алекс	126	75	8	Алита	5	42	
Александр	25	90	1	Алихаджиев	11	39	
Александр	9539		1063	алия	21	85	9
		26790		алкать	18	28	2
Александра	59	238	6	алкаш	31	16	2
александрийский	17	54	1	алкоголизм	87	112	12
Александринка	40	28	2	алкоголик	133	108	22
александринский	49	62	5	алкоголь	488	476	53
Александрия	27	54	2	алкогольный	498	696	30
Александро-невский	21	45		Алла	485	985	62
Александров	181	704	7	Аллах	102	270	36
Александрова	5	37		аллегорический	10	32	
Александрович	596		14	аллегория	14	86	2
		1230		аллегро	3	15	
Александровна	124	216		Аллен	115	156	13
александровский	119	130	7	аллерген	15	8	
Алексашенко	100	227	5	аллергический	37	10	
Алексеев	132	331	18	аллергия	94	84	8
Алексеевич	274	702	9	аллея	110	146	12
Алексеевна	86	131	1	аллигатор	32	17	6
алексеевский	6	26		Аллилуева	14	30	
Алексей	2439		137	аллилуйя	7	19	1
		5687		аллитерация	1	20	
Алексиевич	26	33	3	алло	49	13	25
Алексий	308	852	80	аллюзия	44	126	
Алексин	2	123		Алма-ата	398	764	116
Ален	155	220	23	алма-атинский	31	132	6
Алена	21	147	1	алмаз	697		141
Аленичев	74		3		1428		
Алескеров	1	30		алмазно-бриллиантовый	15	43	2

	А	В	С		А	В	С
алмазный	182	610	63	Альмодовар	4	78	7
алмазо-Бриллиантовый	6	44	1	альпийский	44	111	10
Алмазов	19	74	1	альпинизм	59	61	8
алмазодобывающий	21	57		альпинист	117	97	14
алмазодобытчик	4	16	1	альпинистский	19	33	
алмазодобыча	6	18		альплагерь	12	14	
алмазоносный	2	18	1	Альпы	50	92	24
Алмазювелирэкспорт	18	40		альт	9	140	
алогичность	3	14		альтернатива	344		122
алогичный	5	18	1		1373		
Алпатов	7	32	1	альтернативность	8	38	
Алроса	181	317	36	альтернативный	450		87
алтаец	5	33	2		1328		
Алтай	136	268	46	альтист	11	50	3
алтайский	157	232	25	Альтман	10	36	
алтарный	11	17		альтовый	2	14	
алтарь	93	202	15	альтруизм	21	45	1
алтуфьевский	18	7	1	альтруист	11	19	4
алтын	10	14		альтруистический	10	19	
алуштинский	2	14		Альф	1	46	
алфавит	54	115	6	альфа	254	222	29
алфавитный	22	54	1	Альфа-арт	3	44	2
Алферов	19	84	6	Альфа-банк	212	259	12
Алкан-юрт		43		Альфа-групп	63	56	4
алхимик	12	34	1	Альфонтс	30	17	2
алхимический	1	13		Альфред	417	548	38
алхимия	18	56	3	альянс	899		232
алчность	26	60			4743		
алчный	37	45	3	алюминиевый	343	771	49
алый	73	138	16	алюминий	120	587	24
Альмов	36	4		алюминщик	7	31	5
аль	54	104	6	аляповатый	11	18	
аль-Акса	2	30	1	Аляска	140	212	31
аль-Файед	78	21	5	АМ	16	39	1
альбатрос	4	14	1	Амадей	6	45	3
Альбац	8	33		амазонка	4	20	3
Альбер	11	48	2	Амазонка	20	47	11
Альберт	268	972	34	Амалия	14	31	1
Альберто	75	116	9	Аман	294	699	56
Альбина	27	35		Амангельды	3	70	1
Альбион	86	47	9	Аманда	42	18	2
альбом	1187		36	Амато	7	64	4
		791		амбар	12	26	5
альбомный	1	13		амбарный	22	12	1
Альбрехт	13	27		амбивалентность	3	25	
Альгирдас	86	95	10	амбивалентный	3	25	
Альдо	12	25	1	амбициозность	22	60	1
Альенде	13	67	4	амбициозный	123	341	6
альма-матер	20	13	5	амбиция	410		96
альманах	87	209	11		1487		