Корпус русского рассказа начала XX века. Пример лингвостатистического анализа

А.О. Гребенников, Н.М. Марусенко

Санкт-Петербургский государственный университет

a.grebennikov@spbu.ru, n.marusenko@spbu.ru

Аннотация

Исследование строится на базе представительного «Корпуса русских рассказов 1900 – 1930-х гг.». Для выборки из первого периода включенных в Корпус текстов (100 рассказов с 1900 по 1913 гг. отобранных по принципу не более одного рассказа от каждого из включённых в Корпус авторов) был построен частотный словарь.
С целью выявления ключевых слов, соответствующих основной тематике рассказов, а также прослеживания влияния крупномасштабных политических изменений на язык художественной прозы первые 100 наиболее частотных знаменательных слов словаря были сопоставлены с данными полученных авторами ранее частотных словарей отдельных русских писателей – признанных мастеров рассказа и материалами частотного словаря русского языка в целом. Также было проведено сравнение с данными для русских рассказов аналогичного периода начала XXI века, полученными из НКРЯ. Для получения объективных результатов сравнения использовался показатель числа употреблений на миллион слов (ipm). Полученные результаты показывают, что распределение частот знаменательных слов в верхней зоне словаря может служить хорошим индикатором общей тематики произведений не только отдельного писателя, но и отдельной эпохи, а также отражать актуальные внешние аспекты жизни общества.

Ключевые слова: корпус текстов, русский рассказ, стилеметрия, частотный словарь, распределение частот

Библиографическая ссылка: Гребенников А.О., Марусенко Н.М. Корпус русского рассказа начала XX века. Пример лингвостатистического анализа // Компьютерная лингвистика и вычислительные онтологии. Выпуск 4 (Труды XXIII Международной объединенной научной конференции «Интернет и современное общество», IMS-2020, Санкт-Петербург, 17 – 20 июня 2020 г. Сборник научных статей). — СПб.: Университет ИТМО, 2020. С. 21-28. DOI: 10.17586/0000-0000-2020-4-21-28

Введение

Переломные моменты истории, время общественно-политических перемен, социальных революций неизменно привлекает внимание представителей гуманитарных наук. Их исследования обычно базируются на официальных документах, исторических свидетельствах, публицистике; иногда также привлекаются личные дневники, бытовые записи, частная переписка. Однако ценным источником информации в эпоху кардинальных исторических сдвигов может служить и художественная литература, чутко реагирующая на происходящее вокруг. Особенно это справедливо в отношении жанра рассказа, в котором (в силу его небольшого объема и, следовательно, короткого издательского цикла) оперативно отражаются актуальные события в социальной, политической и культурной жизни, а также проявляются изменения в употреблении языка.

1. Корпус русских рассказов 1900 – 1930-х гг.

Масштаб произошедших языковых сдвигов возможно оценить только с опорой на репрезентативный объем текстов и применение количественных методов обработки материала. С этой целью в СПбГУ реализуется проект по созданию Корпуса рассказов русских писателей, охватывающего произведения возможно большего числа литераторов, написанные на русском языке с 1900 по 1930 гг. и опубликованные в периодических изданиях или отдельными брошюрами (подробнее о принципах построения корпуса см. [1, 2]). Представительность корпуса обеспечивается охватом практически всех литературных направлений и максимального числа творивших в то время литераторов: не только ведущих, но и множества второстепенных. Художественное наследие последних позволяет расширить представления ученых как о разных сторонах общественной и культурной жизни, так и о характерных особенностях языка того времени. На настоящий момент корпус насчитывает несколько тысяч единиц.

Историческим центром указанного периода является Октябрьская революция. Все остальные события и процессы рассматриваются или как преддверие перелома, или как его последствия. В связи с этим корпус разделен на три временных среза [3]: 1) начало XX века и предреволюционные годы, включая Первую мировую войну, 2) революционные годы — Февральская и Октябрьская революции и Гражданская война и 3) постреволюционные годы — с окончания Гражданской войны до 1930 г.

Важность исследования рассказов, относящихся к первому из этих периодов, обусловлена не только серьезными общественно-политическими событиями той поры, но и высказывавшимися предположениями о том, что стилистические изменения могут несколько опережать наступление кризиса и тем самым его прогнозировать [4]. Анализ корпусных данных начинается именно с рассказов, написанных в период 1900-1913 гг. Из них была сформирована выборка, включающая 100 рассказов без каких-либо ограничений по длине, тематике и т. п. (не более чем по одному произведению одного и того же автора) и служащая начальным полигоном для разностороннего изучения материала и выдвижения гипотез, которые в дальнейшем будут проверяться на более обширном массиве текстов. Она содержит рассказы многих широко известных писателей, а именно: Л.Н. Толстого, М. Горького, А.П. Чехова, И.А. Бунина, А.И. Куприна, Л.Н. Андреева, А. Белого, В.Я. Брюсова, К.Д. Бальмонта, В.Г. Короленко, Б.К. Зайцева, А.С. Серафимовича, Д.Н. Мамина-Сибиряка, А.Т. Аверченко, Н. Тэффи. Однако число имен, знакомых разве что специалисту по русской литературе этого периода, заметно преобладает.

Важный аспект изучения стилистических особенностей того или иного произведения – анализ его лексического своеобразия: именно лексика формирует то, что лингвисты называют языковой картиной мира. При совокупном рассмотрении репрезентативного набора текстов (будь то одного автора, одного жанра и т. п.) большое значение имеет частотное распределение слов, причем это касается прежде всего знаменательных частей речи (существительных, прилагательных, числительных, глаголов, наречий), поскольку употребление служебных лексем в целом не отличается разнообразием.

Таким образом, в настоящем исследовании предпринята попытка проанализировать частотное ранжирование знаменательной лексики на материале выборки из 100 рассказов, принадлежащих перу различных авторов и датированных началом XX века. Представляется возможным сравнить верхнюю зону знаменательной лексики с соответствующими зонами писательских словарей и словаря языка в целом. Следует отметить, что исследование подобного рода на представительном материале группы текстов одного жанра и одной эпохи проводится впервые.

2. Методология исследования

На сегодняшний день авторами создан ряд частотных словарей, отражающих особенности авторской стилистики, а именно словари А.П. Чехова, Л.Н. Андреева, А.И. Куприна, И.А. Бунина [5 – 8]. Их материалы позволяют анализировать особенности мировосприятия того или иного писателя, выявлять ключевые темы его творчества.

Несомненный интерес представляет сопоставление верхних зон частотных словарей разных писателей (т. е. наиболее частотных единиц знаменательных частей речи). В статье [9] приведена таблица, содержащая по 50 наиболее частотных знаменательных слов, извлеченных из словарей А.П. Чехова, Л.Н. Андреева, А.И. Куприна, И.А. Бунина. Даже этот, достаточно ограниченный, материал является весьма показательным. Нетрудно заметить, к примеру, что творчество И.А. Бунина отмечено большим числом лексических единиц, относящихся к миру природы (ветер, небо, поле, сад, море, лес, солнце, дорога), а в произведениях А.П. Чехова основное внимание уделяется миру людей (жена, муж, должный, нужный, доктор, слово). Трагическое мироощущение Л.Н. Андреева выражается в высоких рангах слов смерть, черный, темный, плакать [9]. Индивидуально-авторский характер подобных единиц подтверждается также их сопоставлением с материалами частотного словаря русского языка в целом [10].

На материале нашей выборки был также построен частотный словарь общим объемом 24 511 лексем, 383 430 словоформ, позволяющий нам провести сравнение 100 верхних единиц с 1) аналогичными 100 словами из словарей А.П. Чехова, Л.Н. Андреева, А.И. Куприна, И.А. Бунина и 2) аналогичными 100 словами из словаря языка в целом («Частотный словарь русского языка» Шарова – Ляшевской [10]). При этом наше внимание обращено прежде всего на лексические единицы из нашей выборки, отсутствующие в списках, составленных по материалам других словарей (что, разумеется, не означает, что их вовсе нет в других словарях, просто они имеют более низкий ранг и не попали в верхнюю зону частотного распределения). Следует отметить, что ввиду различных объёмов выборок, используемых при создании анализируемых словарей, здесь и далее традиционно используется показатель не абсолютной частоты, а частотности – т.е., числа употреблений на миллион слов (ipm).

3. Результаты

Результаты проведенного исследования выглядят следующим образом. В ходе первого сравнения были обнаружены следующие слова, вошедшие в 100 наиболее частотных знаменательных слов нашей выборки, но отсутствующие у всех четырех писателей: толпа, дети, уйти, никто, тело. Высокую частотность первых двух из них можно объяснить характерной тематикой рассказов тех лет. Так, вхождение в список лексемы дети, вероятно, связано с тем, что многие рассказы изначально создавались для публикации в периодических изданиях и были адресованы широкой аудитории – отсюда преобладание тематики, связанной с браком и семьей. Показательным является высокий ранг слова толпа; можно предположить, что он обусловлен обострением социальных отношений, всплеском политических выступлений, которые нашли отражение в литературе тех лет. Присутствие в списке слов никто, тело и уйти (кстати, видовая пара уходить также не представлена в верхних рангах словарей писателей) едва ли является значимым.

В целом, частотный словарь, созданный на материале рассматриваемой выборки, выглядит гораздо более сбалансированным, чем словари отдельных писателей: индивидуально-авторские черты в нем нивелированы. В итоге некоторые его фрагменты непосредственно отражают смысловые связи между лексическими единицами, чего не наблюдается в словарях языка того или иного писателя. Так, в полученном частотном распределении непосредственно соседствуют слова видеть и смотреть, черный и темный, человек и люди, окно, комната и дверь; одинаковые ранги имеют слова ходить и улица; поблизости друг от друга расположены глаз, рука, лицо и голова.

Второе сопоставление дало гораздо более значительные расхождения. Это неудивительно, поскольку частотный словарь [10] построен на материалах Национального корпуса русского языка (НКРЯ), охватывающего широкий круг речевых жанров современного русского языка. Таким образом, сопоставляемые данные различаются и эпохой, и типом текста, что обусловливает высокий процент несовпадений (45%) в верхних рангах знаменательной лексики. В целом, можно сказать, что частотное упорядочение лексем в словаре современного русского языка отражает актуальные внешние аспекты жизни общества (ср. российский, русский, проблема, вопрос, мир, страна, работа, деньги, машина, иметь, являться, система), в то время как русские рассказы начала XX века более сосредоточены на отдельном человеке – его мыслях, чувствах, отношениях с другими людьми (ср. душа, сердце, чувствовать, бог, мысль, молчать, хотеться, тело, молодой, старый, жена, мать, старик).

Таблица 1. Сопоставление частотности выявленных в анализируемых словарях лексем (число употреблений на миллион слов, ipm). Лексемы, характерные для русского рассказа начала XX века (1 - Корпус русских рассказов (1900-1913); 2 – Русские рассказы (2000 – 2019) в НКРЯ; 3 - Частотный словарь современного русского языка; 4 – Частотный словарь рассказов А.П. Чехова; 5- Частотный словарь рассказов И.А. Бунина; 6 - Частотный словарь рассказов Л.Н. Андреева; 7- Частотный словарь рассказов А.И. Куприна)

Лексема	1	2	3	4	5	6	7
душа	**874**	**126**	**374**	873	487	751	777
сердце	**806**	**370**	**245**	328	742	654	763
молчать	**785**	**399**	**395**	621	414	923	461
толпа	**762**	**121**	**97**	222	280	365	337
жена	**749**	**619**	**377**	1338	426	659	531
дети	**741**	**-***	**-**	520	73	193	350
чувствовать	**735**	**350**	**390**	651	876	619	638
молодой	**717**	**484**	**442**	985	1143	583	742
уйти	**709**	**481**	**316**	495	559	467	368
старый	**707**	**572**	**497**	525	487	563	760
мысль	**702**	**419**	**427**	601	523	898	614
никто	**623**	**860**	**562**	424	450	862	538
старик	**610**	**457**	**395**	596	645	249	496
Бог	**558**	**517**	**301**	1237	1241	462	787
мать	**550**	**352**	**330**	727	414	563	323
хотеться	**545**	**406**	**406**	1141	316	548	236
тело	**542**	**429**	**286**	358	632	969	770

*- лексема не представлена в анализируемой зоне.

Возникает вопрос, какой из двух вышеупомянутых факторов обусловливает это различие. Едва ли на него можно дать абсолютно однозначный ответ. При этом стоит заметить, что далеко не все рассказы посвящены личной жизни отдельных персонажей: как отмечалось выше, драматические события начала XX века нашли отражение в целом ряде произведений. Однако эти события предстают перед читателем в индивидуальном преломлении, через призму автора или персонажа. Это принципиально отличается от способа подачи тех же событий в публицистике, удельный вес которой в НКРЯ довольно высок. Это дает основание предположить, что выявленное различие скорее связано с разницей в характере материала (разные типы текстов в НКРЯ vs. отдельный литературный жанр рассказа в нашем корпусе), чем со сменой эпох. Данная гипотеза может быть проверена на материале некоторого сопоставимого корпуса современных русских рассказов.

Для этого возможно воспользоваться инструментарием НКРЯ и задать поиск выделенных лексем в подкорпусе русских рассказов с 2000 по 2019 год, т.е. практически аналогичном нашей исследуемой выборке. Полученный подкорпус составил 576 документов общим объемом 2 598 738 словоупотреблений, т. о., его представительность не взывает сомнений.

Полученные данные по частотностям выявленных групп лексем в анализируемых словарях (расположенные в порядке убывания частот по данным «Корпуса русского рассказа начала XX века») приводятся в таблицах 1 - 2.

Таблица 2. Сопоставление частотности выявленных в анализируемых словарях лексем (число употреблений на миллион слов, ipm). Лексемы, характерные для современного русского рассказа (1 - Корпус русских рассказов (1900-1913); 2 – Русские рассказы (2000 – 2019) в НКРЯ; 3 - Частотный словарь современного русского языка; 4 – Частотный словарь рассказов А. П. Чехова; 5- Частотный словарь рассказов И. А. Бунина; 6 - Частотный словарь рассказов Л. Н. Андреева; 7- Частотный словарь рассказов А. И. Куприна)

Лексема	**1**	**2**	**3**	**4**	**5**	**6**	**7**
работа	**417**	**649**	**1058**	172	462	269	278
иметь	**389**	**499**	**907**	899	462	527	347
мир	**355**	**595**	**715**	187	389	619	302
российский	**-***	**59**	**645**	20	-	-	7
русский	**-**	**386**	**530**	121	304	36	604
проблема	**-**	**124**	**480**	-	-	-	-
вопрос	**-**	**364**	**806**	318	49	309	198
страна	**-**	**237**	**725**	25	109	30	80
деньги	**-**	**529**	**513**	555	207	314	493
машина	**-**	**543**	**492**	20	97	66	76
являться	**-**	**80**	**523**	45	24	137	69
система	**-**	**66**	**618**	76	-	10	80

*- лексема не представлена в анализируемой зоне.

Наибольший интерес представляют для нас первые три столбца данных таблицы (выделены жирным шрифтом). Легко заметить, что во всех трех выделенных нами группах лексем частотности в корпусах рассказов значительно различаются (как правило, в два и более раз), что позволяет нам утверждать справедливость сделанных предположений, касающихся способности полученных данных выявить лексическое своеобразие эпохи, обусловленное общественно-политической атмосферой и отражающее тенденции в языковом употреблении. Одновременно, последние четыре столбца таблицы еще раз демонстрируют значительную индивидуально-авторскую вариативность выделенных лексем, подтверждая тем самым большой потенциал корпусных и частотных данных в исследовании вопросов индивидуально-авторской стилистики и стилистики вообще, и предоставляя большой простор для дальнейшего содержательного анализа.

4. Обсуждение

В данной статье описываются результаты лингвостатистического анализа частотного словаря, поостренного для выборки объемом 100 рассказов из Корпуса русских рассказов 1900 – 1930-х гг. Исследование затрагивает только первый из рассматриваемых в Корпусе периодов, а именно 1900 – 1913 гг. 100 наиболее частотных знаменательных слов словаря были сопоставлены с данными полученных авторами ранее частотных словарей рассказов отдельных выдающихся русских писателей, а также с материалами частотного словаря русского языка в целом. Далее, было проведено сравнение с данными для русских рассказов аналогичного периода начала XXI века, полученными из НКРЯ.

Для получения объективных результатов сравнения использовался показатель числа употреблений на миллион слов (ipm).

Показано, что рассматриваемые частотные распределения может служить хорошим индикатором общей тематики произведений не только отдельного автора, но и отдельной эпохи в целом. Используемая методика также позволяет проследить влияние крупномасштабных политических изменений на словарный став языка художественной литературы.

Заключение

Полученные результаты следует рассматривать как сугубо предварительные. Дальнейшие исследования будут посвящены аналогичному анализу следующих двух периодов текстов в Корпусе и, затем, всего Корпуса в целом.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант № 17-29-09173 «Русский язык на рубеже радикальных исторических перемен: исследование языка и стиля предреволюционной, революционной и постреволюционной художественной прозы методами математической и компьютерной лингвистики (на материале русского рассказа)».

Литература

[1] Мартыненко Г.Я., Шерстинова Т.Ю., Мельник А.Г., Попова Т.И. Методологические проблемы создания Компьютерной антологии русского рассказа как языкового ресурса для исследования языка и стиля русской художественной прозы в эпоху революционных перемен (первой трети XX века) // Компьютерная лингвистика и вычислительные онтологии. Вып. 2 (Труды XXI Межд. объед. конф. «Интернет и современное общество, IMS-2018, Санкт-Петербург, 30 мая — 2 июня 2018 г.»). СПб.: Университет ИТМО, 2018. С. 99 – 104.

[2] Мартыненко Г.Я., Шерстинова Т.Ю., Попова Т. И., Мельник А. Г., Замирайлова Е. В. О принципах создания корпуса русского рассказа первой трети XX века // Труды XV Международной конференции по компьютерной и когнитивной лингвистике «TEL 2018». Казань: Издательство АН РТ, 2018. С. 180 – 197.

[3] Martynenko G., Sherstinova T. Linguistic and Stylistic Parameters for the Study of Literary Language in the Corpus of Russian Short Stories of the First Third of the 20th Century. In R. Piotrowski's Readings in Language Engineering and Applied Linguistics, Proc. of the III International Conference on Language Engineering and Applied Linguistics (PRLEAL-2019), Saint Petersburg, Russia, November 27, 2019, CEUR Workshop Proceedings. 2020. Vol. 2552. P. 105 – 120.

[4] Lasswell H.D. Style in the language of politics. // Language of Politics: Studies in Quantitative Semantics, 2nd ed., Cambridge, MA, M.I.T. Press. 1965. P. 20 – 39.

[5] Частотный словарь рассказов А.П. Чехова / Авт.-сост. А. О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: СПбГУ, 1999. 120 с.

[6] Частотный словарь рассказов Л.Н. Андреева / Авт.-сост. А. О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: СПбГУ, 2003. 396 с.

[7] Частотный словарь рассказов А.И. Куприна / Авт.-сост. А. О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: СПбГУ, 2006. 552 с.

[8] Частотный словарь рассказов И.А. Бунина / Авт.-сост. А. О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: СПбГУ, 2011. 294 с.

[9] Гребенников А.О. Индивидуально-авторский характер различных зон распределения в частотных словарях языка писателя // Структурная и прикладная лингвистика. Выпуск 11: Межвузовский сборник. СПб.: Изд-во С.-Петерб. ун-та, 2015. С. 100 – 110.

[10] Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2012. 1087 с.

The Early XX-century Russian Short Stories Corpora. An Example of Lingvo-statistical analysis

A.O. Grebennikov, N.M. Marusenko

St. Petersburg State University

The research draws on a representative sample form The Corpus of the Russian Short Stories (1900-1930). A frequency word list was made for the sample from the first group of text in the Corpus (100 stories written between 1900 and 1913, one story per author).
To identify topic words of the stories and the reveal the effect of dramatic events of the period in question on the fiction vocabulary the first 100 most frequent content words in the list are compared with the existing frequency word lists of prominent Russian writers and with the Frequency Dictionary of Russian by Lyashevskaya and Sharov. In addition, the data is set against a representative sample of modern Russian short stories from Russian National Corpus. The conventional measure of relative frequency - instances per million (ipm) - is used to ensure the adequacy of the results. The frequency distribution of content words in the upper zone of the lists is shown to be an indicator of the general short-stories content not only for an individual author, but for the time period as a whole as well as to reflect some actual aspects of public life.

Keywords: text corpus, Russian short stories, stylometry, frequency dictionary, frequency distribution

Reference for citation: Grebennikov A.O., Marusenko N.M. The Early XX-century Russian Short Stories Corpora. An Example of Lingvo-statistical analysis. // Computer Linguistics and Computing Ontologies. Vol. 4 (Proceedings of the XXIII International Joint Scientific Conference «Internet and Modern Society», IMS-2020, St. Petersburg, June 17-20, 2020). - St. Petersburg: ITMO University, 2020. P. 21 – 28. DOI: 10.17586/0000-0000-2020-4-21-28

References

[1] Martynenko G.Y., Melnik A.G., Popova T.I., Sherstinova T.Y. Methodological problems of creating the Computer Anthology of Russian short stories as a language resource designed to study language and style of Russian prose in the era of revolutionary changes (in the first third of the 20th century) // Komp. lingvistika i vychislitel'nye ontologii. Vyp. 2, IMS-2018, SPb: ITMO). 2018. P. 99 – 104. [In Russian].

[2] Martynenko G.Ya., Sherstinova T.Yu., Popova T.I., Melnik A.G., Zamirajlova E.V. O printsipakh sozdaniya korpusa russkogo rasskaza pervoj treti XX veka [On the Principles of Creation of the Russian Short Stories Corpus of the First Third of the XX Century] // Mezhdunarodnaya konferentsiya po komp'yuternoj i kognitivnoj lingvistike "TEL 2018" (TEL 2018. Proceedings of International Conference on Computer and Cognitive Linguistics). Kazan: Izdatelstvo AN PT, 2018. P. 180 – 197. [In Russian].

[3] Martynenko G., Sherstinova T. Linguistic and Stylistic Parameters for the Study of Literary Language in the Corpus of Russian Short Stories of the First Third of the 20th Century // R. Piotrowski's Readings in Language Engineering and Applied Linguistics, Proc. of the III International Conference on Language Engineering and Applied Linguistics (PRLEAL-2019), Saint Petersburg, Russia, November 27, 2019, CEUR Workshop Proceedings. Vol. 2552. 2020. P. 105 – 120.

[4] Lasswell H.D. Style in the language of politics. Language of Politics: Studies in Quantitative Semantics, 2nd ed., Cambridge, MA, M.I.T. Press. 1965. P. 20 – 39.

[5] Grebennikov A.O., Martynenko G.Ya. (Ed.) Chastotnyy slovar rasskazov L. N. Andreeva [Frequency Dictionary of the Short Stories by Leonid N. Andreev]. Saint Petersburg, Izd-vo S.-Peterb. un-ta. 2003. [In Russian].
[6] Grebennikov A.O., Martynenko G.Ya. (Ed.) Chastotnyy slovar rasskazov A. I. Kuprina [Frequency Dictionary of the Short Stories by Alexander I. Kuprin]. SPb: Izd-vo S.-Peterb. un-ta, 2006. [In Russian].
[7] Grebennikov A.O., Martynenko G.Ya. (Ed.) Chastotnyy slovar rasskazov A. I. Bunina [Frequency Dictionary of the Short Stories by Ivan A. Bunin]. SPb: Izd-vo S.-Peterb. un-ta, 2011. [In Russian].
[8] Grebennikov A.O., Martynenko G.Ya. (Ed.) Chastotnyy slovar rasskazov A. P. Chekhova [Frequency Dictionary of the Short Stories by Anton P. Chekhov]. SPb: Izd-vo S.-Peterb. un-ta. 1999. [In Russian].
[9] Grebennikov A.O. Individual'no-avtorskij harakter razlichnyh zon raspredeleniya v chastotnyh slovaryah yazyka pisatelya [Author-specific character of distribution zones in author's frequency dictionaries] // Strukturnaya i prikladnaya lingvistika (Structural and Applied Linguistics), № 11. SPb: Izd-vo S.-Peterb. un-ta, 2015. P. 100 – 110. [In Russian].
[10] Lyashevskaya O.N., Sharov S.A. Chastotnyy slovar sovremennogo russkogo yazyka (na materialakh Natsionalnogo korpusa russkogo yazyka) [Frequency Dictionary of Contemporary Russian based on the RNC Data]. Moscow: Azbukovnik, 2009. [In Russian].