УДК 821.161.1.09:81'38

ПОНЯТИЕ МАРКЕМЫ И ПРЕДВАРИТЕЛЬНЫЕ ИТОГИ МАРКЕМНОГО АНАЛИЗА РУССКОЙ ЛИТЕРАТУРЫ[1]

А. А. Фаустов, А. А. Кретов

Воронежский государственный университет

Поступила в редакцию 20 сентября 2017 г.

Аннотация: *в статье излагаются принципы маркемологического анализа и предпринимается сжатый обзор результатов, полученных в ходе маркемного исследования русской литературы (по преимуществу XVIII – начала XX в.). В основе алгоритма выявления маркем лежит идея, согласно которой наиболее важные для текста словоформы должны употребляться в нем не просто часто, но чаще нормы частоты, характерной для словоформ такой же длины. Маркемные исследования проводились по трем главным направлениям. Во-первых, это изучение в диахронии маркемного состава русской литературы, различных ее эпох и различных текстовых фракций (проза, поэзия). Во-вторых, это определение корреляций: между авторами различных эпох и между авторами и эпохами – диахронический, генеалогический аспект; между авторами одной эпохи или всего периода в целом – синхронический аспект. В-третьих, это исследование синтагматики маркем, явления текстовой аттракции – способности слова притягивать к себе в тексте (на некотором его отрезке) другие слова, синтаксически с этим словом непосредственно не связанные.*

Ключевые слова: *маркема, индекс текстуальной маркированности, текст, слово, частота, русская литература, эпоха, корреляция, текстовый аттрактор, фракция текстов.*

Abstract: *the paper outlines the principles of markeme analysis and presents an overview of the research results obtained in the markemological study of Russian literature (predominantly belonging to XVIII – early XX centuries). The algorithm of the markeme analysis is based on the idea that not only should the words which are the most important for a particular text occur frequently, but with the frequency, which is higher than that of the words of the same length. Markeme research was conducted in three directions. Firstly, the markeme repertoire was studied from a diachronic perspective and in different types of texts (prose vs. poetry). Secondly, correlations between different authors belonging to the same period of time and between authors and various epochs were established, and this is the diachronic genealogical aspect. Thirdly, the research of markeme syntagmatic behavior allowed us to describe the phenomenon of textual attraction, i.e. the ability of the word to attract other, syntactically unrelated words to a certain part of the text.*

Key words: *markeme, index of textual markedness, text, word, frequency, Russian literature, epoch, correlation, textual attractor, fraction of texts.*

Желание объективировать и даже формализовать содержательный анализ текстов (в том числе и в первую очередь – литературных) возникло у филологов отнюдь не вчера. Тем не менее задача эта по сей день не решена, и попытки приблизиться к ее решению не прекращаются.

Цель настоящей работы – представить, во-первых, обоснование еще одного способа такого формализованного изучения текстов, а во-вторых, некоторые уже полученные на этом пути результаты.

Первоначально (особенно интенсивно – в 1960-е гг.) большие надежды возлагались на частотные словари. Однако чем больше их появлялось и чем больше их анализировали, тем больше возникало сомнений в их адекватности, тем яснее становилось, что частота слова в тексте не простая величина, а равнодействующая двух закономерностей – языковой и текстовой.

У каждой словоформы в тексте есть параметр, над которым не властен автор: это ее языковая форма – в том числе и длина в буквах или звуках (русская орфография дает возможность формализовать переход от букв к звукам – пересчитать длину словоформ в звуках).

Таким образом, каждая словоформа в тексте обладает двумя параметрами: частотой и длиной. Каждая из этих величин может быть взвешена: соотнесена с долей слов, обладающих аналогичным, большим или меньшим параметром.

Если частота словоформы является сложным (субъективно-объективным, текстово-языковым) показателем, а длина словоформы – простым (объективным, языковым), то субъективный (содержатель-

[1] Статья выполнена в рамках проекта «Лингвистическая подготовка и создание электронного “Живого стилистического словаря русского языка”», поддержанного грантом РФФИ № 17-04-00421.

© Фаустов А. А., Кретов А. А., 2017

ный, идущий от текста) фактор может быть получен вычитанием объективного фактора (веса словоформы по длине) из субъективно-объективного (веса словоформы по частоте). Полученная величина – Индекс Текстуальной Маркированности словоформы (ИнТеМ) – и будет указывать на степень субъективной (текстовой) весомости данной словоформы для данного текста.

Однако ранжировать словоформы по ИнТеМу недостаточно.

Содержательная составляющая текста также является не простой величиной, а равнодействующей: в первом приближении содержание текста состоит из того, *что* рассказывается (о ком-либо или о чем-либо), и того, *ради чего* рассказывается. И второе явно важнее. При всей условности и относительности такого разделения оно дает критерий (отчасти волюнтаристский) отсеивания лексики, выявленной при вычислении ИнТеМа. Так, в литературном тексте слова, обозначающие персонажей, место, время и обстоятельства их действий, мы должны «вынести за скобки» как относящиеся к тому, *что* рассказывает автор.

При этом мы предполагаем, что оставшиеся слова с их ИнТеМами помогут нам ранжировать слова по их значимости для автора текста и тем самым приблизят нас к ответу на вопрос, *ради чего* текст создан.

Некое множество таких самых важных для автора данного текста словоформ с максимальным значением ИнТеМа мы и называем *маркемами*.

Поскольку понятие маркемы производно от понятия ИнТеМа, предложенного в работах [1; 2], рассмотрим подробнее алгоритм вычисления этой величины.

Как известно, языковая закономерность состоит в большей частоте коротких слов по сравнению с длинными [3; 4]. Чем длиннее слово, тем меньше вероятность его появления в тексте, а чем меньше вероятность его появления в тексте, тем больше информации оно несет. Проще говоря, чем длиннее слово, тем менее оно случайно, тем важнее оно для автора. Понятно, что частотный анализ текста уловить такие редкие, но важные слова не способен.

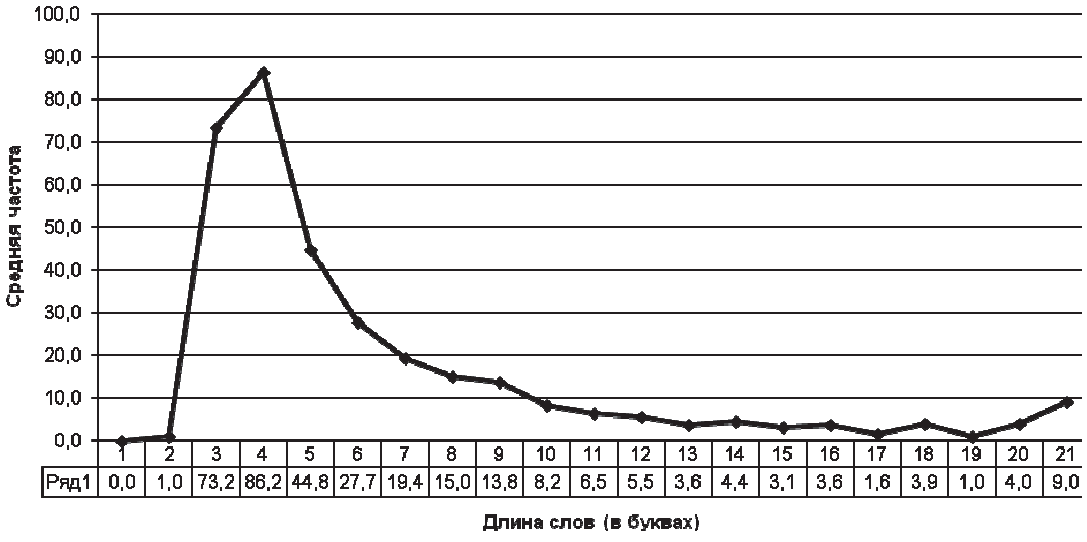
Проверим прежде всего, действительно ли существует зависимость между длиной слов и их частотой в тексте. Эта закономерность представлена на рис. 1 (на примере старославянских текстов).

Как видим, начиная с длины в четыре буквы, т. е. с минимальной нормальной длины полнозначного слова в старославянском языке (три буквы – корень, одна буква – окончание), средняя частота слов данной длины последовательно убывает. Длина же от одной до трех букв – интервал служебных слов. Колебания начинаются там, где кончаются статистически надежные данные.

Рассмотрим на том же материале эту зависимость под иным углом зрения: от частоты к длине. Соответствующая зависимость представлена на рисунке (на примере старославянских текстов) (рис. 2).

Как свидетельствует линейный тренд, по мере убывания частот слов их средняя длина возрастает. Таким образом, данные старославянских текстов подтвердили существование зависимости, обнаруженной Дж. К. Ципфом.

Повторим, что главная идея ИнТеМа проста: для получения текстовой, авторской составляющей из частотной равнодействующей надо вычесть языковую составляющую. Иными словами, словоформа должна



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Ряд1	0,0	1,0	73,2	86,2	44,8	27,7	19,4	15,0	13,8	8,2	6,5	5,5	3,6	4,4	3,1	3,6	1,6	3,9	1,0	4,0	9,0

Рис. 1

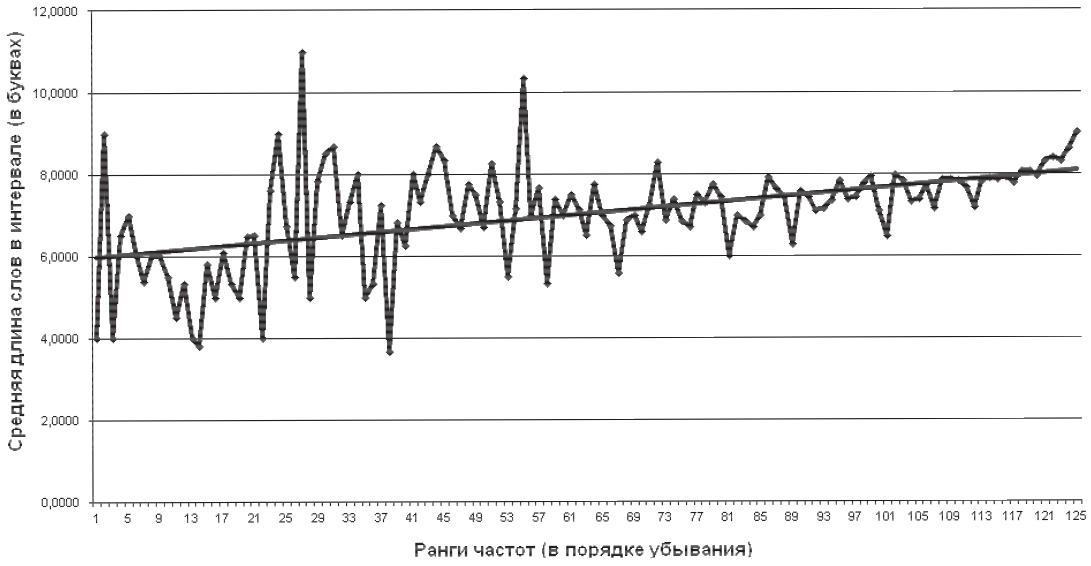


Рис. 2

употребляться в тексте не просто часто, но чаще нормы частоты, характерной для словоформ его длины. Избыточная частота определяет «авторский» вес словоформы, т. е. значение ИнТеМа (подробнее см.: [5]).

Формула вычисления весов [6, с. 15] позволяет привести к одному виду веса словоформ в тексте по частоте и по длине и разместить их в интервале от 1 до 0:

$$Pr_i = \frac{\Sigma r - R_{1-i}}{\Sigma r},$$

где Pr_i – вес единиц i-го ранга, Σr – сумма единиц всех рангов, R_{1-i} – сумма единиц от первого до данного ранга. В данном случае речь идет о рангах частот, а не о рангах слов.

Формула вычисления ИнТеМа:

$$\text{ИнТеМ} = \text{Ч-вес} – \text{Д-вес}.$$

Типичное распределение словоформ по Ч-весу (на примере текстов А. В. Дружинина) представлено на рис. 3.

Типичное распределение словоформ по Д-весу (на примере текстов А. В. Дружинина) представлено на рис. 4.

Получив ИнТеМы словоформ, мы вводим систему фильтров, так или иначе работающих на то, чтобы разделить *что* текста и его *ради чего*:

1. ***Грамматический***. Во-первых, отсеиваются все словоформы, кроме субстантивных. Это объясняется

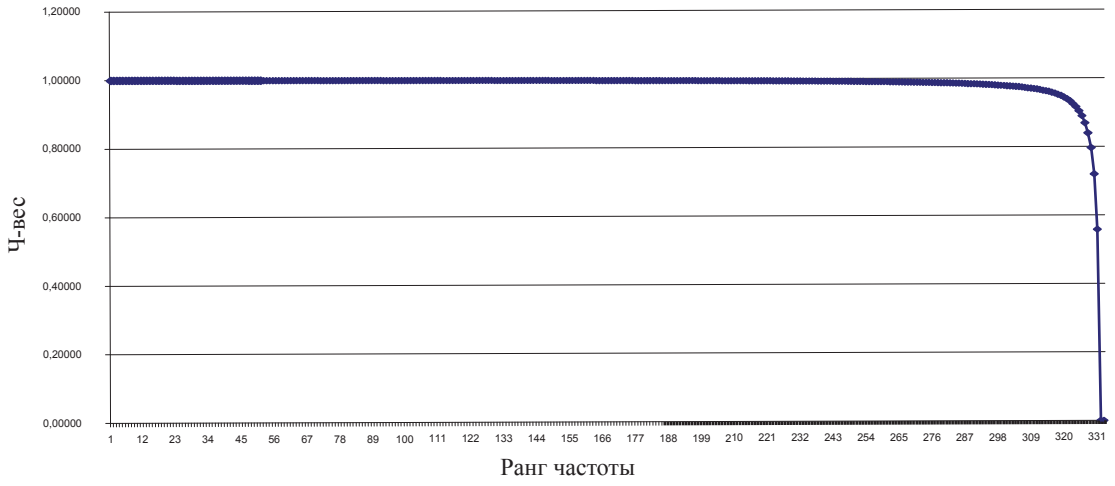


Рис. 3

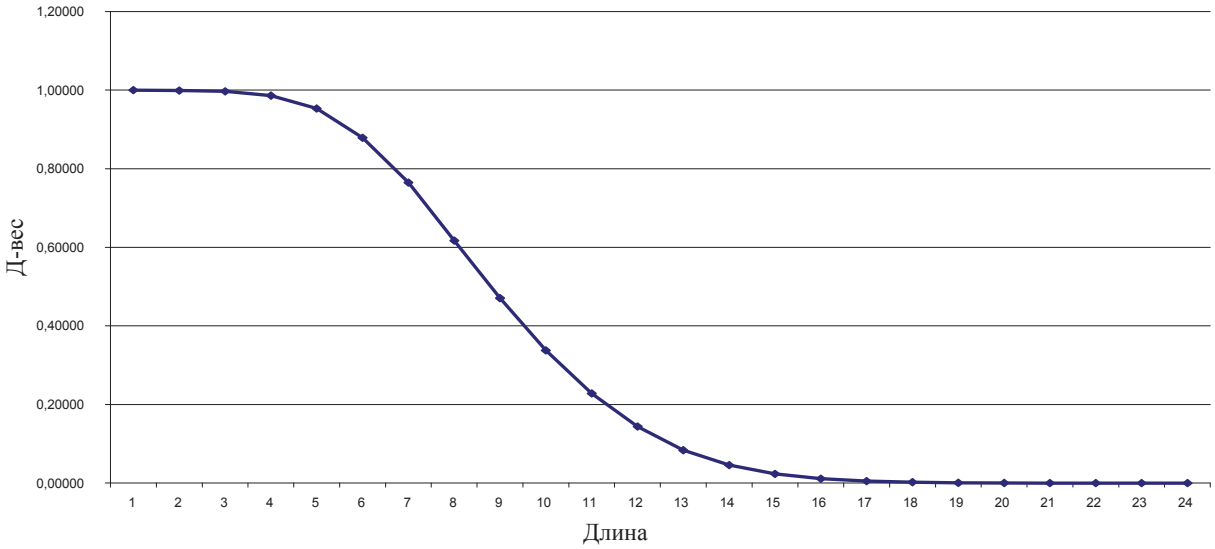


Рис. 4

тем, что существительное является наименее маркированной частью речи (в словаре любого языка существительные преобладают), а кроме того – существительные в максимальной степени ориентированы на внеязыковую действительность. Во-вторых, отсеиваются все субстантивные словоформы, не стоящие в именительном падеже единственного числа (если это не pluralia tantum). Именительный падеж избран в качестве фильтра потому, что он обозначает нечто, способное выступать в качестве субъекта, в том числе сюжетного (если перед нами литературный текст). Кроме того, именительный падеж – это падеж темы высказывания.

2. ***Грамматико-семантический***. Отсеиваются все онимы и одушевленные существительные, кроме словоформы *человек*.

3. ***Тематико-семантический***. Отсеваются существительные, специфические для определенных групп текстов; к примеру: *явление* или *действие* в пьесах; воинская терминология у Д. В. Давыдова и Л. Н. Толстого; названия поэтических жанров; название месяцев и дней недели; название артефактов, если последние не обрели коннотативных / символических значений; и т. д.

4. ***Стилистический***. Отсеиваются стилистически окрашенные слова – диминутивы и т. п.

5. ***Диалогический***. Отсеиваются обращения, звательные формы: *Боже, княже, старче, отче, мати; батюшка, матушка* и т. п.

6. ***Классификаторный***. Отсеиваются слова-классификаторы, фигурирующие в конструкциях «Сущ. И. п. + Сущ. Р. п.», не несущие собственной семантики, как, например, слово *множество* в конструкциях *множество плодов, множество врагов, множество красавиц, множество знамен* и т. д.

С учетом сказанного можно дать такое определение маркемы: ***маркема*** – это одна из 50 словоформ с максимальным значением ИнТеМа, прошедшая через все фильтры (маркем может быть и чуть больше, если у нескольких словоформ одинаковые ИнТеМы). Порог в 50 маркем для одного автора (или литературной эпохи) – величина условная, задаваемая соображениями целесообразности: меньше 50 – мало, больше – много. Имея в виду, что мы сравниваем более 100 авторов попарно друг с другом по 50 маркем, то это будет 12 499 950 сравнений, увеличение же числа маркем вдвое (до 100) потребовало бы в четыре раза больше попарных сравнений – почти 50 млн (49 999 950).

***

Маркемологические исследования русской литературы проводились в основном на материале текстов 112 авторов XVIII – начала XX в. общим объемом более 43 млн словоформ. С некоторой долей хронологической (и не только хронологической) условности этот достаточно замкнутый период был разбит на пять эпох. В основе периодизации лежит культурно-семиотическое понимание поколения, со свойственными ему особой ориентацией во времени, типичным языком переживания реальности, преобладающей формой авторского поведения и т. д. (см.: [7–9]): 1) «классическая» эпоха – созданная писателями, даты рождения которых колеблются – иногда, разумеется, со значительными отклонениями в обе стороны – вокруг 1739 г. (со сгущением в первой половине 1740-х гг.); 2) «элегическая/романтическая» эпоха – созданная писателями, даты рождения которых колеблются вокруг 1796 г. (со сгущением во второй половине 1790-х гг.); 3) «критическая» эпоха

– созданная писателями, даты рождения которых колеблются вокруг 1817 г. (со сгущением в конце 1810-х – первой половине 1820-х гг.); 4) «реалистическая» эпоха – созданная писателями, даты рождения которых колеблются вокруг 1839 г. (со сгущением в 1830-х – начале 1840-х гг.); 5) «модернистская» эпоха – созданная писателями, даты рождения которых колеблются вокруг 1877 г. (со сгущениями во второй половине 1860-х – первой половине 1870-х гг. и вокруг 1890 г.).

Приведем числовые параметры той базы данных, которая послужила основой большинства уже проведенных исследований. Первое число – это номер среза, второе – номер автора, после имени автора – суммарная длина его текстов в словоупотреблениях: 1 1 Богданович 33865; 1 2 Державин 40479; 1 3 Дмитриев И. 57069; 1 4 Кантемир 52561; 1 5 Капнист 21953; 1 6 Карамзин 316203; 1 7 Княжнин 102957; 1 8 Крылов 279194; 1 9 Ломоносов 79938; 1 10 Львов Н. 49905; 1 11 Майков В. 76467; 1 12 Муравьев М. 41174; 1 13 Новиков Н. 241810; 1 14 Радищев 189421; 1 15 Сумароков 83490; 1 16 Тредиаковский 90233; 1 17 Фонвизин 258230; 1 18 Хемницер 43631; 1 19 Херасков 102501; 2 20 Баратынский 54202; 2 21 Батюшков 50738; 2 22 Бестужев-Марлинский 292888; 2 23 Веневитинов 35559; 2 24 Вяземский 60015; 2 25 Гнедич 162232; 2 26 Гоголь 639919; 2 27 Грибоедов 84066; 2 28 Давыдов Д. 83843; 2 29 Дельвиг 51909; 2 30 Жуковский 291327; 2 31 Загоскин 371656; 2 32 Кольцов 23981; 2 33 Кюхельбекер 381712; 2 34 Лермонтов 314899; 2 35 Нарежный 421200; 2 36 Одоевский В.Ф. 367006; 2 37 Погорельский А. 53023; 2 38 Полевой Н. 392658; 2 39 Пушкин 514001; 2 40 Рылеев 49720; 2 41 Языков 78107; 3 42 Аксаков К. 135422; 3 43 Аксаков С. 486978; 3 44 Анненков П. 503854; 3 45 Белинский 1748016; 3 46 Герцен 618756; 3 47 Гончаров 872221; 3 48 Григорович 323381; 3 49 Григорьев А. 476365; 3 50 Дружинин 390540; 3 51 Киреевский И. 36113; 3 52 Майков А. 189006; 3 53 Некрасов 198101; 3 54 Никитин 92194; 3 55 Островский 1074737; 3 56 Писемский 1257313; 3 57 Плещеев 43673; 3 58 Полонский 201417; 3 59 Салтыков-Щедрин 735695; 3 60 Толстой А. 363751; 3 61 Тургенев 1181859; 3 62 Тютчев 61958; 3 63 Фет 205362; 3 64 Хомяков 125286; 4 65 Апухтин 161276; 4 66 Гаршин 74065; 4 67 Добролюбов 680810; 4 68 Достоевский М. 240341; 4 69 Достоевский Ф. 2259643; 4 70 Короленко 358235; 4 71 Леонтьев К. 967100; 4 72 Лесков 1698034; 4 73 Мельников-Печерский 781019; 4 74 Михайловский 425713; 4 75 Надсон 56099; 4 76 Писарев 506226; 4 77 Помяловский 127393; 4 78 Слепцов 98157; 4 79 Случевский 80924; 4 80 Соловьев Вс. 309340; 4 81 Соловьев Вл. 134208; 4 82 Страхов 140058; 4 83 Толстой Л. 3287546; 4 84 Успенский Г. 1142483; 4 85 Успенский Н. 110082; 4 86 Фофанов 35778; 4 87 Чернышевский 568977; 4 88 Чехов 1352918; 4 89 Эртель 341871; 5 90 Андреев Л. 631568; 5 91 Анненский 455519; 5 92 Ахматова 133167; 5 93 Бальмонт 174280; 5 94 Белый А. 300160; 5 95 Блок 543563; 5 96 Брюсов 401542; 5 97 Бунин 770380; 5 98 Волошин 59837; 5 99 Горький М. 431233; 5 100 Гумилев 235243; 5 101 Есенин 113623; 5 102 Иванов Вяч. 462522; 5 103 Кузмин М. 320658; 5 104 Мандельштам 190983; 5 105 Маяковский 423696; 5 106 Мережковский 1024101; 5 107 Пастернак 71750; 5 108 Розанов 497368; 5 109 Сологуб Ф. 464896; 5 110 Ходасевич 555876; 5 111 Цветаева 236315; 5 112 Шмелев 440128.

Тексты каждого автора собирались в единый файл, который обрабатывался специальной программой «ПроТемАЛ» (Программой Тематического Анализа Лексики), вычисляющей ИнТеМы словоформ и ранжирующей их в порядке убывания значений индекса[2]. В результате были выявлены маркемы всего периода в целом, пяти эпох-срезов и каждого из 112 авторов. Ограничимся лишь двумя примерами.

Маркемы русской литературы «классической» эпохи представлены в табл. 1.

Маркемы русской литературы «элегической/романтической» эпохи представлены в табл. 2.

Как видно даже из этих примеров, к маркемам относятся в первую очередь:

1) абстрактная лексика (преимущественно на *-и(ь)е, -ство, -ость/-есть*), обозначающая важнейшие категории культуры: *добродетель, необходимость, человек, счастье, искусство, богатство, блаженство, удовольствие, воображение, вдохновение, впечатление* и т. д.:

2) имена натурфактов и артефактов, выступающих в качестве субъекта действия и нередко наделенных коннотативными/символическими значениями: *солнце, звезда, земля, небо, сердце, лес, берег, река, кинжал* и т. д.

Не претендуя на исчерпывающий характер и абсолютную точность, можно было бы предложить в рабочем порядке такую более развернутую тематическую классификацию маркем (ср. иначе: [10]). Их возможно распределить по девяти группам (которые иллюстрируются некоторыми ключевыми для всего исследуемого периода словами). Во-первых, это «антропологическая» лексика – то, что обозначает относящееся к сущности человека и его социально-биологической конституции (*человек*, *человечество*, *сердце*, *здоровье*). Во-вторых, это «институциональная» лексика – то, что обозначает экономические, социально-политические и культурно-психологиче-

[2] Идея А. А. Кретова, руководство И. Е. Ворониной, исполнение И. Поповой.

Таблица 1

№	Маркема	ИнТеМ
1	Добродетель	8,90093
2	Человек	8,06447
3	Сердце	6,63103
4	Множество	6,48540
5	Богатство	5,97682
6	Счастье	5,65084
7	Блаженство	5,63603
8	Искусство	5,63198
9	Отечество	5,09592
10	Спокойствие	4,39251
11	Природа	4,12278
12	Должность	3,47540
13	Солнце	3,42873
14	Надежда	3,26482
15	Достоинство	3,06254
16	Удовольствие	2,95115
17	Смертный	2,88281
18	Любовь	2,80220
19	Красота	2,71288
20	Воздух	2,66932
21	Премудрость	2,61553
22	Престол	2,54242
23	Источник	2,47186
24	Несчастье	2,42111
25	Невинность	2,41255
26	Приятность	2,39188
27	Праздник	2,33199
28	Справедливость	2,19828
29	Причина	2,17245
30	Чувствительность	2,16374
31	Смерть	2,15891
32	Благодарность	2,12218
33	Храбрость	2,11094
34	Страсть	2,06070
35	Состояние	2,00923
36	Невежество	1,98107
37	Красавица	1,95463
38	Торжество	1,86216
39	Здоровье	1,84139
40	Пространство	1,81960
41	Предмет	1,79605
42	Сладость	1,77709
43	Отчаянье	1,75566
44	Радость	1,72543
45	Благополучие	1,72341
46	Время	1,69369
47	Дружество	1,66029
48	Склонность	1,62337
49	Сокровище	1,59481
50	Воображение	1,57449

Таблица 2

№	Маркема	ИнТеМ
1	Человек	9,79118
2	Воображение	7,91815
3	Сердце	7,49727
4	Мгновение	6,43222
5	Счастье	5,97048
6	Чувство	5,92726
7	Внимание	5,17681
8	Отечество	5,12982
9	Искусство	4,59807
10	Природа	4,38527
11	Добродетель	4,29740
12	Любопытство	4,13421
13	Блаженство	4,13376
14	Молчание	4,02201
15	Достоинство	3,90843
16	Наслаждение	3,90010
17	Спокойствие	3,85591
18	Провидение	3,65092
19	Доказательство	3,63836
20	Благодарность	3,62086
21	Богатство	3,58311
22	Вдохновение	3,31558
23	Радость	3,28559
24	Просвещение	3,19822
25	Воспоминание	3,18185
26	Время	3,06917
27	Солнце	2,98275
28	Впечатление	2,96304
29	Благословение	2,86105
30	Характер	2,63889
31	Возможность	2,57562
32	Удовольствие	2,54407
33	Справедливость	2,52500
34	Страданье	2,42061
35	Состояние	2,39829
36	Движение	2,24676
37	Человечество	2,22807
38	Веселье	2,18416
39	Здоровье	2,18116
40	Обязанность	2,15755
41	Задумчивость	2,15570
42	Самолюбие	2,13336
43	Сладость	2,05747
44	Любовь	1,93011
45	Бессмертие	1,91707
46	Прелесть	1,89801
47	История	1,85806
48	Свобода	1,82016
49	Необходимость	1,75429
50	Отчаяние	1,72926

ские условия и механизмы существования человека (*отечество*, *богатство*, *состояние* – в значении ‘звание’,‘богатство, имение’, *достоинство* – в значении ‘стоимость, ценность’,‘отличие, звание’, *образование* – в значении ‘просвещение’). В-третьих, это «императивная» лексика – то, что обозначает ценности, нормы и правила, регулирующие социальное поведение человека (*добродетель*, *справедливость*, *преступление*, *благодарность*). В-четвертых, это «артистическая» лексика – то, что обозначает принадлежащее к сфере искусства и эстетических отношений человека к миру (*искусство*). В-пятых, это «эмоционально-ментальная» лексика – то, что обозначает реакции субъекта на внутренние и внешние стимулы (*чувство*, *удовольствие*, *счастье*, *беспокойство*). В-шестых, это «коммуникативная» лексика – то, что обозначает принадлежащее к сфере сообщения субъекта с миром, с другими субъектами и с самим собою (*внимание*, *впечатление*, *выражение*, *воображение* – в значении ‘способность ментально представлять те или иные предметы’). В-седьмых, это «ориентационная» лексика – то, что обозначает взаимное соотношение элементов мира и ментальных фактов и отношение их к субъекту, а также субъекта к ним и к другим субъектам (*расположение*, *направление*, *противоположность*, *присутствие* – в значении ‘пребывание в наличии’). В-восьмых, это «космологическая» лексика – то, что обозначает мир в целом, наиболее существенные его элементы и формы его бытия (*природа*, *солнце*, *время*, *пространство*). В-девятых, это «онтологическая» лексика – то, что обозначает фундаментальные, общие структуры бытия (*существование*, *действительность*, *необходимость*, *возможность*).

* * *

По каким направлениям возможно исследовать полученные данные?

Первое их них обусловлено одной особенностью предложенного способа вычисления ИнТеМа (и, соответственно, определения маркем), которая связана с преобладанием Д-веса над Ч-весом. Иначе говоря, в итоге преимущество будет на стороне длинных слов, а поскольку в различных «фракциях» литературных текстов (об этом понятии см.: [11]) средняя длина слов различна (в порядке возрастания ряд таков: драма > поэзия > проза > публицистика) [12], приоритет в смешанных по фракционному составу файлах будет на стороне «публицистических» маркем.

Продемонстрируем это на нескольких примерах. Для начала рассмотрим публицистические (Пу), прозаические (Пр), поэтические (По) и драматические (Др) тексты Н. М. Карамзина и проанализируем расстояние между ними по их маркемам.

Как видим из рис. 5, наименее удаленными друг от друга по составу маркем оказались проза и публицистика. Ярко выраженной спецификой обладают драматические и поэтические тексты.

Репрезентативность фракций текстов по сравнению с совокупной выборкой по данным текстам Н. М. Карамзина убывает в следующей последовательности: поэзия + проза, проза, публицистика, поэзия, драма (рис. 6).

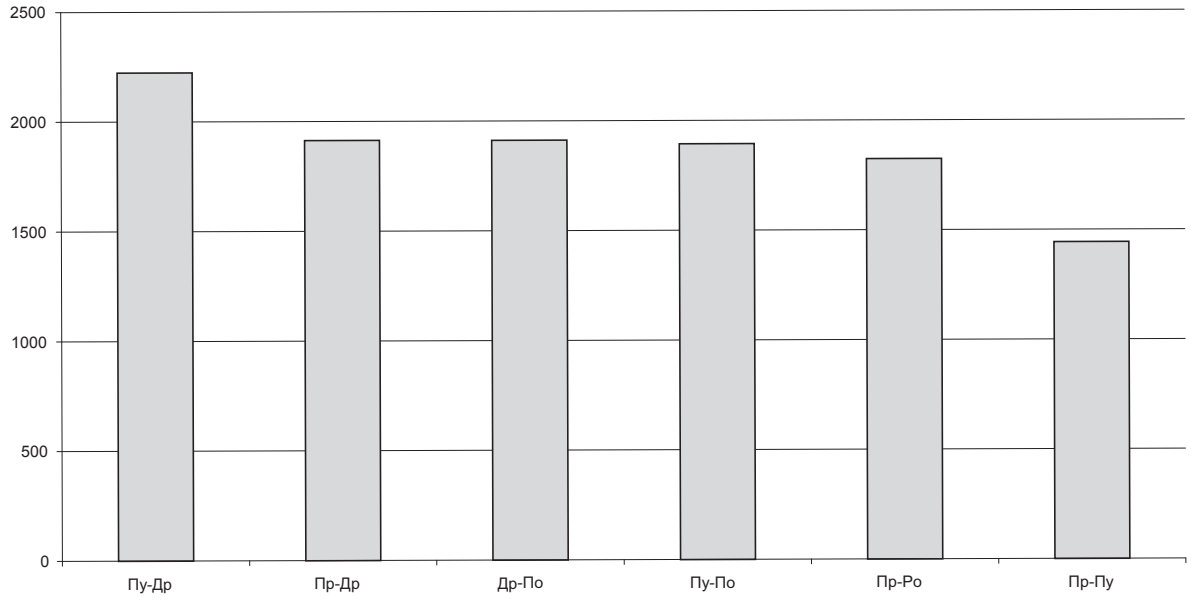


Рис. 5

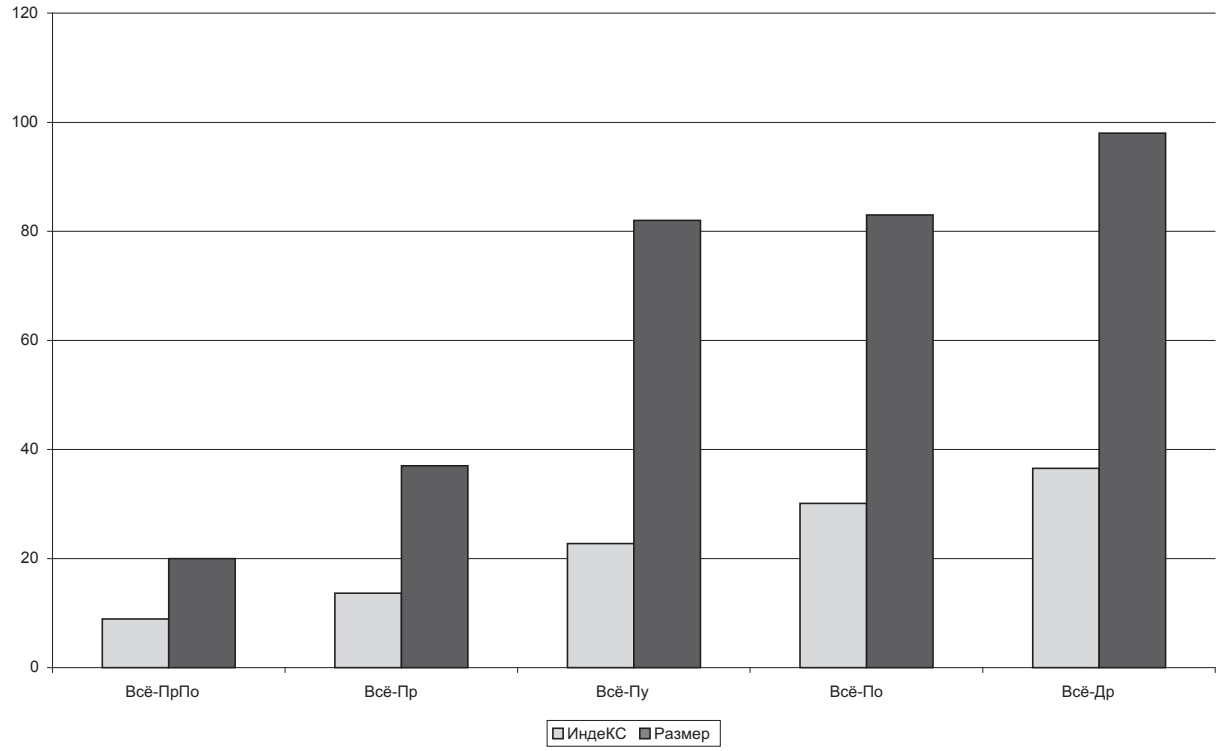


Рис. 6

Влияние фракций на состав маркем в текстах Ф. М. Достоевского и Л. Н. Толстого дает не менее интересную информацию.

Маркемы прозы и публицистики Достоевского совпадают на 27 %: *впечатление; удовольствие; беспокойство; доказательство; воспоминание; справедливость; необходимость; обстоятельство; соображение; человечество; недоразумение; предчувствие; действительность; предположение.*

Специфику прозы Достоевского отражают маркемы: *человек; мгновение; любопытство; положение; сердце; наслаждение; состояние; выражение; характер.*

О специфике публицистики Достоевского говорят маркемы: *самосовершенствование; противоположность; самостоятельность; сантиментальничание.*

Маркемы прозы и публицистики Толстого совпадают на 23 %: *предположение; самоотвержение; несправедливость; справедливость; ответственность; противоположность; представление; подтверждение; действительность; существование; удовлетворение; последовательность.*

Специфику прозы Толстого отражают маркемы: *обстоятельство; воспоминание; неудовольствие; расположение; усовершенствование* и др.

О специфике публицистики Толстого говорят маркемы: *совершенствование; жертвоприношение; священнодействие; распространение; жизнепонимание* и др.

Репрезентативность прозы и публицистики по сравнению с совокупной выборкой по данным текстов Достоевского выглядит следующим образом.

Общие для всех трех множеств (проза, публицистика, проза + публицистика) маркемы составляют 27 %.

Ближе к совокупной выборке оказывается публицистика: маркемы совпадают на 58 %. Проза дальше от совокупной выборки: 31 % общих маркем.

Исключительно между собой маркемы публицистики и прозы совпадают на 29 % маркем, отсутствующих в совокупной выборке.

Маркемы публицистики, как видим, доминируют в совокупной выборке, несмотря на то, что по объему выборка прозы больше выборки публицистики: 83 % художественных текстов против 17 % публицистических.

Если маркемы совокупной выборки специфичны на 38 %, а маркемы публицистики – на 40 %, то маркемы прозы у Достоевского специфичны на 67 % (подробнее о Толстом см.: [13]).

Сопоставление маркем столь разных авторов, как А. Т. Твардовский и О. Э. Мандельштам (ср.: [14]), дает материал для столь же любопытных выводов. Так, поэтические тексты Твардовского и Мандельштама объединяют шесть общих для поэзии начала XX в. (и во многом общих для всей русской поэзии XVIII – начала XXI в.) маркем: *красота; праздник; радость; сердце; солнце; человек*.

Максимальная же маркемная близость Твардовского и Мандельштама на генеральных совокупностях текстов (поэзия + проза + публицистика) обеспечивается суммой прозаических и публицистических текстов. Из 22 маркем Твардовского и Мандельштама, полученных на множестве текстов «поэзия + проза + публицистика», 18 являются общими и для выборки «проза + публицистика»: *расположение; существование; ответственность; противоположность; сопротивление; расстояние; отсутствие; революция; возвращение; отношение; принадлежность; представление; стремление; впечатление; действительность; удовлетворение; обстоятельство; настроение.* Остальные четыре (*воспоминание; выражение; пространство; состояние)* не входят ни в число общих «прозаических + публицистических», ни в число общих «поэтических» маркем этих двух поэтов.

Следовательно, близость маркем, полученных на совокупностях «поэзия + проза + публицистика» Твардовского и Мандельштама, свидетельствует о близости непоэтических текстов авторов и, по-видимому, прежде всего публицистики, ибо доля прозы в наследии обоих относительно невелика, а публицистика, как мы видели на примере Толстого и Достоевского, даже на меньших объемах «перевешивает» прозу.

Все сказанное показывает, что маркемы также являются не простой, а сложной сущностью: среди них можно выделить маркемы, навязываемые автору принадлежностью текстов к той или иной фракции. Следовательно, если мы хотим с помощью маркем проникнуть в мир автора, мы должны учитывать фракционную специфику его текстов. И то же самое справедливо по отношению к литературной эпохе.

Более того, маркемная неоднородность литературного творчества и литературного процесса в качестве особой задачи ставит проблему исследования «маркемной» истории отдельных литературных фракций. И на некоторые данные таких «парциальных» исследований уже можно сослаться. Так, на обширном материале русской поэзии XVIII–XXI вв. под руководством А. А. Кретова была выполнена кандидатская диссертация А. В. Кашкиной «Маркемный анализ языка русской поэзии» [15], результаты которой представлены в серии публикаций [16–22] и в табл. 3, 4, а также дипломные работы студентки 5 курса д/о ОТиПЛ факультета РГФ ВГУ Е. М. Мотовых на тему «Маркемный анализ русской прозы конца XVIII – начала XXI века» (Воронеж, ВГУ, кафедра теоретической и прикладной лингвистики, 2012 г.) и студентки 6 курса в/о ОТиПЛ факультета РГФ ВГУ Л. М. Томаленко на тему «Маркемный анализ языка русской драматургии конца XVIII – начала XX вв.» (Воронеж, 2015).

Таблица 3

Маркемы русской поэзии «классической» эпохи

Маркема	СумИнТеМ
Добродетель	3,78832
Блаженство	3,19515
Спокойство	2,84953
Пространство	2,08647
Счастье	1,95165
Совершенство	1,88992
Богатство	1,84294
Премудрость	1,83716
Человек	1,75838
Отечество	1,51907
Сердце	1,31065
Приятность	1,28065
Невинность	1,26236
Природа	1,24051
Чувствительность	1,12905
Надежда	1,06057
Достоинство	1,03354
Солнце	0,90756
Естество	0,90211
Торжество	0,88191
Воздух	0,83418
Несчастье	0,79325
Радость	0,78885
Небеса	0,74980
Отчаянье	0,73074
Красота	0,72729
Бесконечность	0,71846
Сладость	0,71706
Благородство	0,70132
Храбрость	0,64850
Страсть	0,60408
Притворство	0,59082
Благословенье	0,57448
Благополучие	0,55652
Пламень	0,54239
Свирепость	0,53175
Наслажденье	0,52850
Восторг	0,52443
Молния	0,52432
Здоровье	0,51496
Гордость	0,51151
Злодейство	0,50385
Искренность	0,49596
Дружество	0,47981
Чувство	0,44079
Судьбина	0,43169
Заблужденье	0,42733
Любовь	0,42082
Мужество	0,41764
Время	0,40149

Таблица 4

Маркемы русской поэзии «элегической/романтической» эпохи

Маркема	СумИнТеМ
Воспоминанье	3,78043
Благословенье	2,56331
Наслажденье	2,37202
Блаженство	2,35743
Спокойствие	1,82620
Счастие	1,74229
Воображенье	1,56782
Добродетель	1,53822
Отечество	1,42712
Мгновенье	1,41536
Провиденье	1,20989
Уединенье	1,18979
Человек	1,16571
Бессмертие	1,16261
Сердце	1,11386
Младость	1,11200
Стремленье	1,08562
Очарованье	1,05674
Разговор	1,03146
Преступленье	1,01610
Сновиденье	1,01441
Сладострастье	1,00906
Волненье	0,98420
Природа	0,97768
Радость	0,93678
Заблужденье	0,86010
Богатство	0,85119
Надежда	0,82921
Пространство	0,76142
Небосклон	0,75757
Сладость	0,74840
Веселье	0,73753
Призрак	0,72601
Благодарность	0,68061
Задумчивость	0,66645
Исступление	0,63435
Солнце	0,61379
Прелесть	0,60227
Приговор	0,55653
Пламень	0,55584
Мечтанье	0,51020
Поцелуй	0,50829
Свобода	0,44432
Несчастье	0,43391
Взор	0,41169
Восторг	0,39480
Любовь	0,37958

Фракции текстов имеют свои собственные наборы маркем. Наиболее ярко это проявляется в языке поэзии. Поэтическими маркемами являются: *человек; счастье; сердце; солнце; время; здоровье; праздник; радость; взгляд; красота; молодость; смерть; старость*. Они встречаются у самых разных поэтов только потому, что они поэты и вынуждены пользоваться языком русской поэзии.

Поэтические маркемы с максимальным значением произведения количества авторов на СумИнТеМ: *человек 1196,995; голос 145,1352; сердце 1000,753; богатство 139,518; счастье 885,566; надежда 136,808; солнце 766,884; красота 131,308; мгновение 522,188; ветер 121,953; блаженство 463,995; простор 119,203; воспоминание 424,794; призрак 113,791; пространство 390,571; земля 110,353; радость 357,369; небосклон 107,814; природа 327,247; наслажденье 107,368; воздух 319,030; отечество 104,329; молчание 309,161; звезда 104,015; разговор 301,847; небеса 97,209; вдохновенье 218,692; молодость 94,488; любовь 216,019; свобода 87,175; время 211,752; проклятие 74,551; благословение 197,002; слово 71,157; страдание 191,904; пламя 69,297; добродетель 188,409; чувство 67,739; спокойствие 188,141; стремительность 67,57292; бессмертие 186,9255; восторг 65,26051; взгляд 179,0033; предчувствие 64,443; смерть 153,355; народ 62,811; торжество 153,168; одиночество 60,208; глаза 145,697; праздник 57,364* [15, с. 296].

***

Второе направление в изучении маркемного состава русской литературы связано с определением различного рода корреляций – между авторами различных эпох и между авторами и эпохами – диахронический, генеалогический аспект, между авторами одной эпохи или всего периода в целом – синхронический аспект (подробнее см.: [23–29]).

Первый из «генеалогических» вопросов имеет две стороны, и его можно переформулировать так: 1) за кем следовал тот или иной писатель, 2) кто следовал за ним? (разумеется, речь здесь идет не о влияниях и даже не об интертекстуальности).

Получив «авторские» ряды маркем, возможно определить «коэффициент корреляции», т. е. коэффициент, определяющий то, насколько лингвостатистически близки друг к другу авторы. Предварительно веса маркем нормируются так, чтобы сумма квадратов весов для каждого автора равнялась единице. Это было сделано для того, чтобы при нахождении «коэффициента корреляции» маркемы разных авторов имели сопоставимый вес.

«Коэффициент корреляции» (K_{ij}) получается в результате умножения весов совпадающих маркем и сложения результатов, а поскольку веса нормированы,

значения коэффициентов варьируются от 0 до 1: 0 – абсолютное расхождение (для авторов, у которых нет ни одной общей маркемы), 1 – абсолютное схождение (для корреляции автора с самим собой).

При этом «потомком» признается такой автор из предшествующих временных срезов, у которого наибольшее количество общих с данным автором маркем, а при прочих равных условиях – такой, у которого максимальный из всех авторов-предшественников K_{ij} принадлежит данному автору, т. е. такой, для которого нет (по K_{ij}) автора, ближе данного.

Результаты нашего исследования можно обобщить в виде литературно-генеалогического древа (рис. 7).

Разумеется, к полученным результатам надо подходить с некоторой долей здоровой критичности и осмотрительности. Следует иметь в виду, что эти результаты (как и маркемные характеристики эпох) напрямую зависят, по меньшей мере, от двух факторов: во-первых, от полноты текстов автора, привлеченных к исследованию; во-вторых, от корпуса авторов, для которых устанавливаются лингвостатистические связи. (Более редкие несистемные факторы, по-видимому, вообще не поддаются исчерпывающему учету.) Например, если бы в наш корпус были включены такие авторы, как И. З. Суриков и поэты-суриковцы, то, возможно, они (по используемым параметрам) оказались бы ближе к Есенину, чем Гаршин. Однако расширение корпуса авторов чревато противоположной опасностью: выражаясь в терминах теории связи, мы можем утратить способность отличать сигнал от шума. На такую мысль наводит, в частности, подозрительно обширное литературное «потомство» Всеволода Соловьева.

Второй из «генеалогических» вопросов позволяет приблизиться к проблеме «писатель и эпоха» или, иначе говоря, «писатель и его литературные предшественники, современники, последователи».

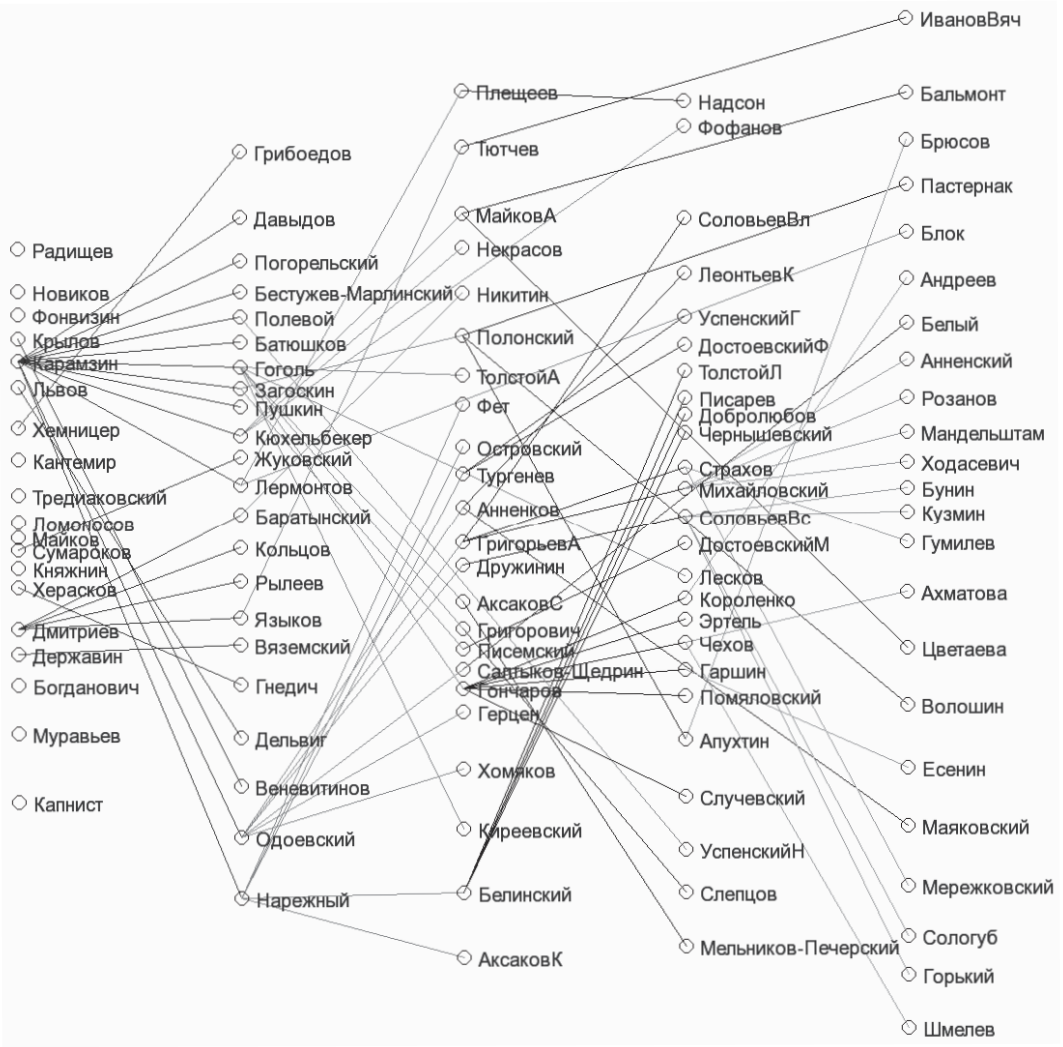


Рис. 7. Маркемно-генеалогическое древо русской литературы XVIII – начала XX в.

Имея таблицу корреляции наборов маркем для всех 112 русских авторов XVIII – начала XX в. (по принципу «каждый с каждым»), мы можем определить отношение каждого из авторов к своему хронологическому срезу, а также – к предшествующим и последующим.

Для этого надо вычислить среднее значение Kij для всех авторов каждого из 5 хронологических срезов по отношению к каждому из 112 авторов.

Если максимальное значение среднего хронологического Kij по отношению к данному автору окажется в одном из предшествующих срезов, автора можно считать «архаистом», если к одному из последующих, то перед нами «новатор» (воспользуемся в своих целях известными тыняновскими терминами), если же максимум среднего хронологического Kij находится в срезе, к которому принадлежит автор, то это «актуальщик».

Из 112 авторов больше всего актуальщиков – 66, в два раза меньше архаистов – 33 и почти в три раза меньше, чем архаистов, новаторов – 13.

По срезам динамика архаистов – новаторов – актуальщиков распределяется следующим образом (в относительных величинах) (По вполне понятным причинам данных по архаистам «классической» эпохи и новаторам «модернистской» эпохи у нас нет.):

	Срез 1	Срез 2	Срез 3	Срез 4	Срез 5
Актуальщики	0,95	0,73	0,43	0,40	0,52
Архаисты	–	0,18	0,17	0,56	0,48
Новаторы	0,05	0,09	0,39	0,04	–

«Авторские» корреляции могут быть рассмотрены и в синхроническом аспекте, что дает возможность установить степень близости автора к своим «соседям» по срезу, а также выявить определенные авторские кластеры.

На основе полученных для всех авторов «коэффициентов корреляции» возможно создавать «карты русской литературы», позволяющие как раз более наглядно представить полученные данные[3]. Для этого необходимо превратить «коэффициенты корреляции», определяющие близость авторов, в расстояния, определяющие удаленность авторов друг от друга. Если аналогичным образом нормировать и эти расстояния ($R_{ij} = 1 – K_{ij}^2$), то расстояние также будет изменяться в диапазоне от 0 до 1: 0 – расстояние от автора до абсолютно такого же автора или до него самого, 1 – расстояние между абсолютно непохожими авторами, K_{ij} которых был нулевым.

При построении карты необходимо нанести на плоскость множество авторов – точек – так, чтобы расстояние между каждыми двумя было бы равно рассчитанным величинам. А поскольку это математически недостижимо (карта должна была бы быть многомерной), задачу можно упростить: необходимо нанести на плоскость множество авторов, расстояние между которыми было бы больше или равно рассчитанным величинам.

Когда все авторы нанесены на плоскость в виде точек, мы получаем «карту» того или иного хронологического среза или сводную «карту русской литературы». Приводим для примера карты двух срезов (рис. 8 и 9).

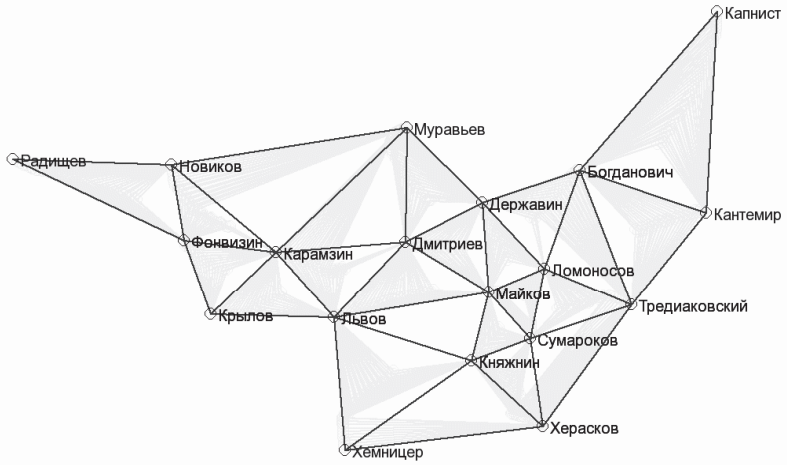


Рис. 8. Карта русской литературы «классической» эпохи

[3] Алгоритм построения карты разработан М. В. Катовым.

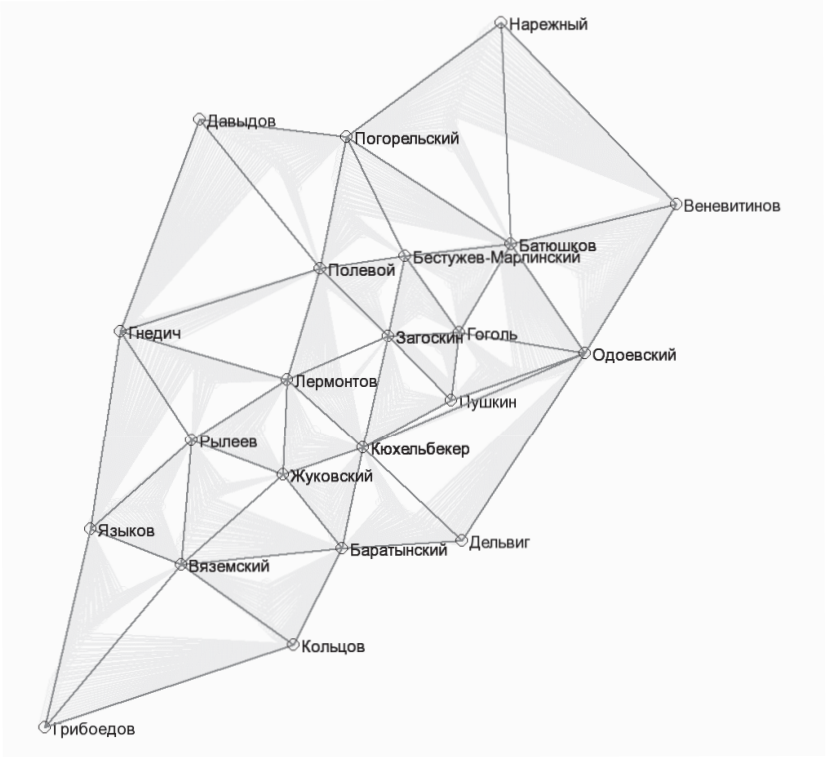


Рис. 9. Карта русской литературы «элегической/романтической» эпохи

***

Наконец, третье направление маркемологических разысканий – это изучение синтагматического поведения маркем и его сопоставление с поведением других, «обычных» слов (подробнее см.: [30]).

По аналогии с лексической аттракцией, т. е. повышенной валентностью слова, его способностью вступать в сочетания с другими словами, здесь можно было бы говорить о более широком явлении «текстовой аттракции» – способности слова как бы притягивать к себе в тексте (на определенном его отрезке) другие слова, синтаксически с этим словом непосредственно не связанные, но образующие с ним своего рода «сверхсочетания».

Объективная оценка текстовой аттракции стала возможной благодаря созданной М. В. Катовым программе «Konstellations», позволяющей количественно определять «коэффициент констелляции» слов (K_k) – степень подобия в распределении любых двух слов в тексте – и графически представлять результаты подобных расчетов.

После расчета коэффициентов констелляции для всех пар маркем из этого ряда выбирается 25 пар слов с наибольшим значением коэффициента, и они изображаются в виде графа, вершины которого – слова, а ребра – линии их «сверхсочетаний». Такого рода слова и предлагается именовать «текстовыми аттракторами»; при этом аттрактивная способность слова тем выше, чем больше степень вершины, т. е. количество исходящих из вершины ребер. Построение системы аттракторов различной силы и различного ранга позволяет внести дополнительные смысловые акценты в выявленные комплекты авторских, «эпохальных», общелитературных маркем.

Кроме того, программа «Konstellations» дает возможность определять набор текстов, в которых маркемы присутствуют с наибольшей густотой. Такие тексты могут рассматриваться как ключевые (или прецедентные) для автора, эпохи, всего изучаемого периода русской литературы и служить предметом особенно пристального исследовательского внимания. (Из числа весьма эффектных результатов сошлемся на то, что для Чехова ключевым произведением оказывается «Скучная история», причем одновременно она является в чеховском творчестве и тем текстом, в котором с наибольшей густотой присутствуют маркемы как «реалистической» эпохи, так и всего исследуемого периода русской литературы.)

Приведем два графа, отражающих активность поведения маркем в текстах двух эпох-срезов (рис. 10 и 11).

***

Наконец, в заключение нужно сказать, по меньшей мере, об одной обнаружившейся проблеме, связанной с семантизацией маркем.

Проблема эта имеет две стороны.

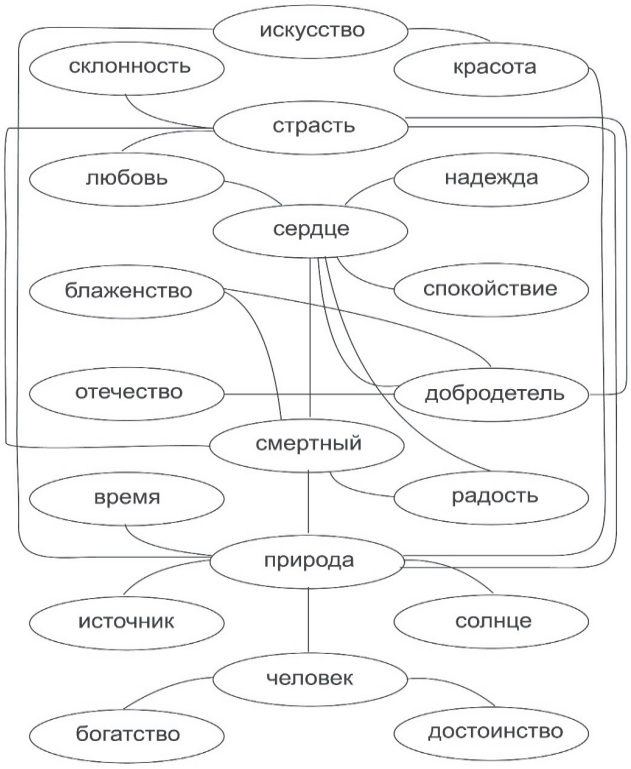


Рис. 10. Текстовые аттракторы «классической» эпохи

Рис. 11. Текстовые аттракторы «элегической/романтической» эпохи

Одна – полисемичность подавляющего большинства маркем, спектр, веса и динамику значений которых возможно уяснить, только обратившись к текстовому массиву, а это, учитывая объем базы данных, задача «вручную» почти не выполнимая.

Укажем хотя бы на маркему *человек*, которая занимает первое место в ряду общелитературных маркем, доминирует в двух эпохах-срезах и у 20 авторов. Так вот, достаточно беглого просмотра текстов исследуемого периода, чтобы убедиться, что *человек* этот далеко не всегда (а вернее, очень редко) – та самая «тайна», которую разгадывал Достоевский. Едва ли не по большей части *человек* – или счетная единица в выражениях вроде «тысяча человек», или отвлеченный, почти десубстантивированный индекс в таких словосочетаниях, как «молодой человек», «богатый человек», «честный человек» и т. д. (огромное разнообразие подобных формул дают, к примеру, произведения Загоскина), или слово со значением ‘слуга’. В силу этого делать напрямую по ИнТеМам маркемы какие-либо слишком далеко идущие выводы о смысле соотношений между авторами и/или эпохами оказывается рискованным.

Вторая сторона проблемы – историко-литературная интерпретируемость полученных данных. Как показывает опыт истолкования чеховского комплекта маркем [31], задачи подобного рода являются необычайно трудоемкими технически и несколько гадательными по решению (последнее, впрочем, можно скачать о любом истолковании). Но дело не только и не столько в этом. Значительная часть результатов (причем это касается всех направлений маркемных исследований) с точки зрения привычных представлений оказывается неожиданной, а иногда и вызывающей. Если, скажем, «Скучная история» в роли ключевого для Чехова (а отчасти и для всей русской литературы) текста выглядит вполне убедительно, то воспринимать в таком же качестве официальную записку «О народном воспитании» у Пушкина или рецензию «Записки ружейного охотника Оренбургской губернии…» у Тургенева сложно.

Одним словом, результаты маркемологических разысканий нуждаются в интерпретационном ключе, и поиск его – задача на будущее.

ЛИТЕРАТУРА

1. *Кретов А. А.* Метод формального выделения тематически нейтральной лексики (на примере старославянских текстов) / А. А. Кретов // Вестник Воронеж. гос. ун-та. Сер.: Системный анализ и информационные технологии. – 2007. – № 1. – С. 81–90.

2. *Кретов А. А.* Опыт выявления архетипов поэзии А. В. Кольцова / А. А. Кретов // Лінгвістичні студії : Зб. наук. праць. Вип. 16 / укл.: Анатолій Загнітко (наук. ред.) [та ін.]. – Донецьк : ДонНУ, 2008. – С. 353–366.

3. *Zipf G. K.* Human Behavior and the Principle of Least Effort / G. K. Zipf. – Cambridge : Addison-Wesley Press, 1949.

4. *Zipf G. K.* The Psycho-Biology of Language an introduction to dynamic philology / G. K. Zipf. – Cambridge : Mass. MIT Press, 1965.

5. *Кретов А. А.* Понятие маркемы : методика выявления и практика использования / А. А. Кретов // Универсалии русской литературы 2 : сб. статей / под ред. А. А. Фаустова ; Воронеж. гос. ун-т. – Воронеж : Наука-Юнипресс, 2010. – С. 138–153.

6. *Титов В. Т.* Частная квантитативная лексикология романских языков / В. Т. Титов. – Воронеж : Изд-во Воронеж. гос. ун-та, 2004.

7. *Фаустов А. А.* Авторское поведение в русской литературе : середина XIX века и на подступах к ней / А. А. Фаустов. – Воронеж : Изд-во Воронеж. гос. ун-та, 1997.

8. *Фаустов А. А.* Язык переживания русской литературы : на пути к середине XIX века / А. А. Фаустов. – Воронеж : Воронеж. гос. ун-т, 1998.

9. *Фаустов А. А.* Игры воображения : историческая семантика характера в русской литературе / А. А. Фаустов, С. В. Савинков. – Воронеж : Научная книга, 2013.

10. *Кашкина А. В.* О классификации маркем / А. В. Кашкина // Вестник Воронеж. гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация. – 2011. – № 1. – С. 97–100.

11. *Фаустов А. А.* О фракциях литературных текстов : к обоснованию понятия / А. А. Фаустов // Вестник Воронеж. гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация. – 2012. – № 2. – С. 34–38.

12. *Кретов А. А.* Роль средней длины словоформы в маркемном анализе / А. А. Кретов // Универсалии русской литературы. 3 / под ред. А. А. Фаустова. – Воронеж : Научная книга, 2011. – С. 16–27.

13. *Кретов А. А.* Л. Н. Толстой : публицист или художник? / А. А. Кретов // Филологические записки : Вестник литературоведения и языкознания. Вып. 30. – Воронеж : Воронеж. гос. ун-т, 2010–2011. – С. 78–98.

14. *Кретов А. А.* Наследие А. Т. Твардовского в диалоге с прошлым / А. А. Кретов // Наследие А. Т. Твардовского в диалоге с прошлым, настоящим и будущим : сб. науч. статей, посвящ. 100-летию со дня рождения А. Т. Твардовского / под ред. В. М. Акаткина. – Воронеж : Наука-Юнипресс, 2010. – С. 415–431.

15. *Кашкина А. В.* Маркемный анализ русской поэзии : дис. … канд. филол. наук / А. В. Кашкина. – Воронеж, 2013.

16. *Кашкина А. В.* Критерии выделения маркем в русских поэтических текстах / А. В. Кашкина // Среди нехоженых путей : сб. науч. статей к юбилею д-ра филол- наук, проф. А. А. Кретова / под ред. И. А. Меркуловой, К. М. Шилихиной. – Воронеж : Наука-Юнипресс, 2012. – С. 300–307.

17. *Кашкина А. В.* Маркемная специфика русской поэзии 60–80-х годов XX века / А. В. Кашкина // Вестник Воронеж. гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация. – 2013. – № 2. – С. 90–94.

18. *Кашкина А. В.* Маркемный состав языка русской поэзии первой трети XX века // Мир лингвистики и коммуникации : электронный научный журнал. – 2011. – № 1 (22). – Режим доступа: http://tverlingua.ru/archive/022/5_22.pdf

19. *Кашкина А. В.* Особенности маркемной лексики языка русской поэзии начала XX века / А. В. Кашкина // Русское национальное сознание в его языковом воплощении : прошлое, настоящее, будущее. XXX Распоповские чтения : материалы Междунар. конф. (Воронеж, 2–4 марта 2012 г.). – Воронеж : Изд.-полиграф. центр Воронеж. гос. ун-та. – 2012. – С. 103–106.

20. *Кашкина А. В.* Русская поэзия 20–50-х годов XX века в ее маркемологическом отображении / А. В. Кашкина // Вестник Воронеж. гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация. – 2013. – № 1. – С. 79–87.

21. *Кашкина А. В.* Язык русской поэзии конца XIX в. с точки зрения маркемологии / А. В. Кашкина // Всемирный виртуальный конгресс по русистике и культуре «Планета «Русский язык» в виртуальном лингвокоммуникативном пространстве» (3–5 октября 2012 г.) / Лингвистический центр Болонского университета CLI, Форли, Италия. – Ереван : Limush, 2012. – Режим доступа: http://community.middlebury.edu/~beyer/MESIV/WVCongress2012.pdf ; http://mesi.cliro.unibo.it/data/_uploaded/file/atti/WVCongress.pdf – С. 254–260.

22. *Кашкина А. В.* XVIII век в русской поэзии с точки зрения маркемного анализа / А. В. Кашкина // Образовательные технологии в виртуальном лингво-коммуникативном пространстве : IV Междунар. виртуал. Конф. по русистике, литературе и культуре : сб. науч. докладов. – Ереван : Лимуш, 2011. – С. 109–113.

23. *Кретов А. А.* Архаисты и новаторы в русской литературе XVIII – начала XX вв. / А. А. Кретов // Универсалии русской литературы / под ред. А. А. Фаустова. – Воронеж : Изд. дом Алейниковых, 2009. – С. 29–48.

24. *Кретов А. А.* И. А. Гончаров на фоне русской литературы (маркемологический анализ) / А. А. Кретов // Универсалии русской литературы. 4 / под ред. А. А. Фаустова. – Воронеж : Научная книга, 2012. – С. 123–147.

25. *Кретов А. А.* Поэзия А. В. Кольцова в контексте русской литературы XVIII – начала XX вв. / А. А. Кретов, М. В. Катов // А. В. Кольцов вчера, сегодня, завтра : материалы Межвуз. науч. конф. / Воронеж. гос. ун-т. – Воронеж : Изд. дом Алейниковых, 2009. – С. 137–150.

26. *Кретов А. А.* Сквозь призму маркем : Н. В. Гоголь в ближайшем контексте русской литературы / А. А. Кретов, М. В. Катов // Вестник Воронеж. гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация. – 2009. – № 2. – С. 12–21.

27. *Кретов А. А.* Кто же вышел из «Шинели» Гоголя? (Опыт лингвостатистической генеалогии) / А. А. Кретов,

М. В. Катов, А. А. Фаустов // Материалы по русско-славянскому языкознанию : Междунар. сб. науч. трудов. – Вып. 30. – Воронеж : ИПЦ ВГУ, 2010. – С. 45–52.

28. *Кретов А. А*. Лингвостатистическая генеалогия в русской литературе XVIII – начала XX вв. / А. А. Кретов, М. В. Катов, А. А. Фаустов // Проблемы компьютерной лингвистики : сб. науч. трудов / под ред. А. А. Кретова. – Воронеж, 2010. – Вып. 4. – С. 114–125.

29. *Кретов А. А*. Опыт лингвистической генеалогии (на примере Н. В. Гоголя) / А. А. Кретов, М. В. Катов, А. А. Фаустов // Проблемы изучения живого русского слова на рубеже тысячелетий : материалы V Всерос. науч.-практ. конф. / науч. ред. А. Д. Черенкова. – Воронеж : ВГПУ, 2009. – С. 76–83.

30. *Фаустов А. А.* Творчество А. П. Чехова в свете явления «текстовой аттракции» и ключевые произведения русской литературы / А. А. Фаустов, М. В. Катов, А. В. Гостева // Универсалии русской литературы. 2 / под ред. А. А. Фаустова. – Воронеж : Наука-Юнипресс, 2010. – С. 154–179.

31. *Фаустов А. А*. Маркемный портрет А. П. Чехова / А. А. Фаустов // Вестник Воронеж. гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация. – 2010. – № 2. – С. 12–16.

Воронежский государственный университет

Фаустов А. А., доктор филологических наук, профессор, заведующий кафедрой русской литературы XIX века

E-mail: aafaustov@list.ru

Тел.: 8 (473) 220-84-98

Кретов А. А., доктор филологических наук, заведующий кафедрой теоретической и прикладной лингвистики

E-mail: a_a_kretov@rambler.ru

Тел.: 8 (473) 220-41-49

Voronezh State University

Faustov A. A., Doctor of Philology, Professor, Head of the Russian Literature of 19th Century Department

E-mail: aafaustov@list.ru

Tel.: 8 (473) 220-84-98

Kretov A. A., Doctor of Philology, Head of the Theoretical and Applied Linguistics Department

E-mail: a_a_kretov@rambler.ru

Tel.: 8 (473) 220-41-49