

**ИГОРЬ ПЕШКОВ**

(Москва, Россия)

ORCID 0000-0001-9881-3429

e-mail: ivpeshkov@gmail.com

**ПОВТОРНОЕ В НЕПОВТОРИМОМ,  
ИЛИ КАК ПРОСЧИТАТЬ ИНДИВИДУАЛЬНЫЙ СТИЛЬ  
(Статистический эпилог проблемы «Бахтин под маской»)**

**REPEATING IN UNREPEATABLE OR HOW TO MEASURE  
INDIVIDUAL STYLE  
(The Statistical Epilogue of the issue *Bakhtin behind the Mask*)**

**Abstract:**

In the article the author focuses on the method of attribution of text on the basis of writer's style, i.e. statistics of repeated word combinations. At first the algorithm was applied to the text corpus of Russian literature, then it was used to analyze the so-called disputed works of Mikhail Bakhtin. Language is a set of words and principles of their compatibility, while the text (or speech in the oral texture of use of language) is the very combination of words in all types of human communications. Accordingly, the style is a special combination of words in the text. To capture the unique style of a writer in a network of statistical analysis it is necessary to quantify the repeated elements of the style. The author calculated repeated two-word combinations (bigrams) of compared texts. A certain quantitative level of repeated bigrams in different texts indicates that these texts were written by the same author. This level varies for different types of texts. The author considered two bodies of literary texts and two bodies of works of humanities. The result of this analysis clearly indicates that the books "Formal Method in Literary Scholarship" and "Marxism and the Philosophy of Language" were

written by the same person, i.e. Mikhail Bakhtin. Additionally, with the highest degree of probability the author attributes the monograph "Freudianism" to Bakhtin as well.

**Keywords:** Mikhail Bakhtin, statistics, style, disputed works of the Bakhtin circle, attribution

Одним из пионеров математических методов поиска в тексте чего-то аналогичного отпечатку пальца биографического автора был физик-метеоролог из США Т.К. Менденхолл. Он предположил идентифицировать автора по частотному распределению в тексте слов разной длины<sup>1</sup>. Это предположение не подтвердилось, но с тех пор статистические методики определения авторства публикуются регулярно. Исследования часто проводятся для решения проблем атрибуции культурно значимых произведений, таких как шекспировские пьесы и поэмы. За этой научной областью с 1890-х годов закрепилось название „стилеметрия“<sup>2</sup>, ее библиография насчитывает не одну сотню единиц.

В России стилеметрический подход заработал с начала XX века, когда Н.А. Морозов высказал мысль, что средством определения индивидуальности стиля может быть распределение в тексте служебных слов<sup>3</sup>. Свои подходы к количественному определению авторства предложили М.А. Марусенко<sup>4</sup>, А.А. Поликарпов<sup>5</sup>, Д.В. Хмелёв<sup>6</sup>.

---

<sup>1</sup> Mendenhall: *The characteristic curves of composition*, „Science“ 1887, 9 (214), p.237-249.

<sup>2</sup> Wincenty Lutoslawski: *Principes de stylométrie appliqués à la chronologie des œuvres de Platon*, „Revue des Études Grecques“ 1898, t. 11, fascicule 41, p.61-81.

<sup>3</sup> Николай Морозов: *Лингвистические спектры: средство для отличения плагиатов от истинных произведений того или иного известного автора. Стилеметрический этюд*, „Известия отд. русского языка и словесности Имп. Акад. Наук“ 1915, т. XX, кн. 4.

<sup>4</sup> Михаил Марусенко: *Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов*. Ленинград 1990.

<sup>5</sup> Ольга Кукушкина, Анатолий Поликарпов, Дмитрий Хмелёв: *Определение авторства текста с использованием буквенной и грамматической информации*, „Проблемы передачи информации“. Москва 2001, т. 37, № 2, с.96-108.

Уже в текущем столетии в сфере цифровых гуманитарных исследований большую популярность приобрел стилеметрический метод Delta, разработанный Дж. Барроузом<sup>7</sup>. Только за последние несколько лет появились десятки работ, развивающих методiku стилеметрического анализа Delta<sup>8</sup>. Известно, что Delta хорошо работает и с русским языком<sup>9</sup>. В целом Delta в первом приближении решает задачу определения авторства спорных текстов количественными методами, хотя есть некоторые частные проблемы, как статистико-ма-

<sup>6</sup> Дмитрий Хмелёв: *Распознавание автора текста с использованием цепей* А.А. Маркова, „Вестник МГУ“ 2000, сер. 9: Филология, № 2, с.115-126.

<sup>7</sup> John Burrows: *'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship*, „Literary and Linguistic Computing“ 1 September 2002, Volume 17, Issue 3, p.267-287, <https://doi.org/10.1093/lc/17.3.267> (19.02.2021)

<sup>8</sup> Вот далеко не полный перечень: Alexander Gladwin, Matthew Lavin, Daniel Look: *Stylometry and collaborative authorship: Eddy, Lovecraft, and 'The Loved Dead'*, „Digital Scholarship in the Humanities“ 2017, v. 32, issue 1, 1 April, p.123-140, <https://doi.org/10.1093/lc/fqv026> (19.02.2021); José Tello: *What does Delta see inside the Author?: Evaluating Stylometric Clusters with Literary Metadata*, in: *III Congreso de La Sociedad Internacional Humanidades Digitales Hispánicas: Sociedades, Políticas, Saberes*. Málaga: HDH, pp.153-161, Hartmut Ilsemann: *Stylometry approaching Parnassus*, „Digital Scholarship in the Humanities“ 2018, v. 33, issue 3, 1 September, p.548-556, <https://doi.org/10.1093/lc/fqx058> (19.02.2021); Michael Oakes: *Computer stylometry of C. S. Lewis's The Dark Tower and related texts*, „Digital Scholarship in the Humanities“ 2018, v. 33, issue 3, 1 September, p.637-650, <https://doi.org/10.1093/lc/fqx043> (19.02.2021); Jacques Savoy: *Is Starnone really the author behind Ferrante?* „Digital Scholarship in the Humanities“ 2018, v. 33, issue 3, 1 September, p. 902-918, <https://doi.org/10.1093/lc/fqy016> (19.02.2021). Одновременно с этим свою популярность в современной науке стилеметрия подтверждает и другими публикациями, в которых используется не Delta, а другие методы определения авторства: Mark Eisen, Alejandro Ribeiro, Santiago Segarra, Gabriel Egan: *Stylometric analysis of Early Modern period English plays*, „Digital Scholarship in the Humanities“ 2018, v. 33, issue 3, 1 September, p.500-528, <https://doi.org/10.1093/lc/fqx059> (19.02.2021); Andrea Nini: *An authorship analysis of the Jack the Ripper letters*, „Digital Scholarship in the Humanities“, 2018, v. 33, issue 3, 1 September, p.621-636, <https://doi.org/10.1093/lc/fqx065> (19.02.2021).

<sup>9</sup> Даниил Скоринкин, Анастасия Бонч-Осмоловская: „Особые приметы“ в речи художественных персонажей: количественный анализ диалогов в „Войне и мире“ Л. Н. Толстого, „История“, 2016, т. 7, № 7 (51).

тематические (снижение качества при уменьшении числа слов в анализируемом тексте ниже 5000, ошибки в работе с жанрово гетерогенными корпусами текстов), так и теоретико-филологические: недостаточная интерпретируемость связи между понятием стиля и статистическими результатами программы.

Мы коснемся теоретических проблем, поскольку в этом есть необходимость, и попытаемся применить выборочные стилеметрические методы главным образом для решения задач, связанных в первую очередь с атрибуцией спорных текстов М.М. Бахтина и его круга. К таким спорным текстам, как известно, относятся три книги: *Фрейдизм и Марксизм и философия язык*» (далее сокр. *МФЯ*) с титульной атрибуцией В.Н. Волошинову, а также *Формальный метод в литературоведении* (сокр. *ФМЛ*), где на месте автора зафиксирован П.Н. Медведев. Кроме того, к спорным относится ряд статей середины – конца 20-х годов, изданных под именами указанных авторов, а также одна статья, подписанная к печати именем И.И. Канаева. Дискуссии вокруг атрибуции всех этих текстов принято называть бахтинским вопросом – по аналогии с гомеровским вопросом или шекспировским.

Впервые в связи с бахтинским вопросом статистический анализ на ограниченном текстовом материале был применен в самом конце прошлого столетия<sup>10</sup>. Теперь цифровые технологии сделали очередной скачок в развитии и у нас есть возможность уточнить методику анализа и осуществить его на материале всех «спорных текстов» и текстов, бесспорно написанных Бахтиным, с использованием в исследованиях больших фоновых корпусов.

Однако прежде чем погрузиться в стихию конкретных подсчетов, необходимо дать базовые теоретические определения. Будем исходить из того, что любой естественный человеческий язык – это **совокупность слов и принципов их сочетаемости**, а текст (или речь в устной фактуре применения языка) – это само **сочетание слов** во всех видах человеческой коммуникации. Нам эти несколько тривиальные и в принципе слишком общие дефиниции нужны для исходного понимания того, что такое стиль. Стоит ли добавлять

---

<sup>10</sup> Игорь Пешков: *Конец – «Делу» венец (Промежуточный текстологический финиш в бахтинском вопросе, или Еще раз об авторстве М. М. Бахтина в «спорных текстах»)*, „Диалог. Карнавал. Хронотоп“ 2000, № 1, с.72-109.

к стилю эпитет «индивидуальный»? Вроде бы стоит, ведь говорим же мы о стиле группы писателей или ученых, о стилевых особенностях жанров речи и т.д. и т.п., вплоть до стиля эпохи. Однако, это все-таки производное, отчасти метафорическое употребление термина. Первоначально стиль – в полном соответствии с латинской внутренней формой слова – уже предполагает связь с индивидуальностью пишущего или говорящего. Если не нравится чей-то стиль, то, как говорил Маяковский, «вот вам, товарищи, мое стило и можете писать сами».

Так мы приходим к формулировке: **стиль это индивидуально-особенное сочетание слов** в тексте. Как сказал другой поэт, «каждый пишет, как он дышит», но и в дыхании есть то, что присуще всем людям (вдох-выдох), и сугубо индивидуальные черты (например, его частота и глубина). Соответственно, сочетания из двух-трех слов в письменной фиксации крайне редко бывают индивидуально-неповторимы. Но если мы возьмем текст из четырех-пяти-шести или более слов, то комбинаторика усложнится и уже не так просто будет найти аналогичное сочетание слов где бы то ни было еще.

Таким образом, в любом тексте есть повторные элементы и элементы неповторимые, те самые, которые и создают индивидуальность стиля. Возникает вопрос, как же уловить и оценить стиль, если он неповторим? Теоретически очень просто: количественно оценить составляющие его повторные элементы, хотя это простое решение выглядит как парадокс: стиль проявляется в неповторимых кусках текста, а просчитывать мы предлагаем элементы повторяющиеся, прежде всего, все двусловные сочетания (биграммы) определенного текста. Но тем-то язык и его применение как дискретная форма коммуникации отличается от стиля поведения, например, демонстрации эмоций, или других форм коммуникации (в существенной мере континуальных), что **в индивидуальных**, тоже можно сказать в известной степени континуальных проявлениях языка (**высказываниях**), скрываются дискретные, **повторные элементы** (слова языка и их взаимные сочетания).

Если у текста есть типичные черты стиля, обычно заметные тренированным глазом или ухом, они проявятся и через первичные сочетания слов, число которых можно посчитать. Последняя фраза, например, вот так разбивается на эти первичные сочетания слов (биграммы): *если у, у текста, текста есть, есть типичные, типичные черты, черты стиля, стиля обычно, обычно заметные, заметные*

тренированным, тренированным глазом, глазом или, или ухом, ухом они, они проявятся, проявятся и, и через, через первичные, первичные сочетания, сочетаниях слов, слов число, число которых, которых можно, можно посчитать<sup>11</sup>. Некоторые из биграмм очень похожи на традиционные словосочетания, некоторые – не очень и выглядят случайными стыками разных слов. На результатах исследований это не сказывается, так как для методики анализа значимы только повторные сочетания, единичные при сравнении текстов отпадают сами собой.

При сравнении. Это ключевые слова методики: считать биграммы имеет смысл прежде всего в сравнении (в котором, как известно, все и познается). Итак, сравниваем два текста, получаем список одинаковых биграмм (сочетаний двух слов) и их количество. Чем этих **общих биграмм больше, тем тексты больше похожи друг на друга своим стилем**. Такова исходная лемма. И ее придется проверить в первую очередь. Метод проверки эмпирический, но, чтобы было от чего отталкиваться, предполагаем как общепризнанное, что наиболее похожими являются тексты из одного и того же произведения одного автора.

Итак, сначала сравним на наличие одинаковых биграмм разные части одних и тех же произведений одного и того же автора. Затем – разные произведения одного автора. И, наконец, – разные произведения разных авторов.

Обращаясь к исходной лемме (в частности, чтобы окончательно определиться, лемма ли она), возьмем для примера *Записки ружейного охотника* С.Т. Аксакова и сравним между собой части этого произведения (по 15000 слов каждая, всего пять частей). Результат отразился в таблице:

	1	2	3	4	5
1		1211	1166	1172	1177
2	1211		1301	1383	1187
3	1166	1301		1354	1203
4	1172	1383	1354		1364
5	1177	1187	1203	1364	

<sup>11</sup> Количество биграмм определяется по формуле  $V=2n-2$ , где  $n$  – это количество слов в анализируемом тексте.

Верхняя строка и первый слева столбец показывают, какие части сравниваются между собой. Результат (число одинаковых биграмм в сравниваемых текстах) фиксируется в клеточке на пересечении столбца и строки. Поскольку сравнение отрывка самого с собой бессмысленно, то на пересечении тождественных текстов клетка остается пустой, а поскольку все части пересекаются в таблице между собой дважды, то дубли просто не принимаются во внимание (например, можно вообще не смотреть на левый нижний треугольник этой таблицы).

Как видим, для нашего примера одинаковых биграмм в разных парах сравнения обнаруживается от 1166 (минимум) до 1364 (максимум). Среднее арифметическое таких одинаковых биграмм на одну пару сравнения (в данном случае  $[1211+1166+1172+1177+1301+1383+1187+1354+1203+1364] : 10 = 1251,8$ ) можно назвать показателем уровня гомогенности стиля произведения. Или просто уровнем гомогенности.

Теперь сравним *Записки ружейного охотника* (части обозначены в левом столбце) с другими произведениями Аксакова же (также в отрывках по 15000 слов<sup>12</sup>):

	Детские годы Багрова_1	Детские годы Багрова_2	Детские годы Багрова_3	Детские годы Багрова_4	Детские годы Багрова_5	Семейная хроника_1	Семейная хроника_2	Семейная хроника_3	Семейная хроника_4	Записки и обрывки_1	Записки и обрывки_2	
1	706	641	662	656	661	672	634	594	572	586	891	907
2	683	695	674	613	683	660	643	577	561	594	823	1060
3	715	668	704	666	725	717	660	626	607	647	851	979
4	707	693	694	675	702	666	680	595	605	625	837	945
5	674	655	696	688	690	707	676	602	610	580	852	971

Мы видим, что даже самый высокий уровень совпадения биграмм в парах сравнения разных произведений Аксакова (1060) здесь ниже самого низкого уровня при сравнении отрывков одного и того же произведения (*Записки ружейного охотника*): 1166. А среднее арифметическое результатов этих пар сравнения разных произведе-

<sup>12</sup> Ниже, если не определено иное, речь всегда будет идти об анализе отрывков именно такого размера.

ний Аксакова (уровень **гомогенности стиля автора**<sup>13</sup>) вообще будет радикально отличаться от уровня гомогенности стиля произведения: **697** против **1252**.

Еще ниже уровень совпадений биграмм, если сравнивать тексты Аксакова с отрывками произведений других авторов. Например, с произведениями Амфитеатрова:

	Кня жна _1	Кня жна _2	Кня жна _3	Кня жна _4	Отрав ленна я совест ь_1	Отрав ленна я совест ь_2	В стр ане люб ви_1	В стр ане люб ви_2	Жа р- цвете т_1	Жа р- цвете т_2	Жа р- цвете т_3	Жа р- цвете т_4	Жа р- цвете т_5
1	495	469	493	476	494	488	513	506	513	518	558	562	536
2	465	457	467	440	460	459	451	507	537	505	520	531	534
3	456	492	481	466	493	485	502	529	560	539	534	562	506
4	476	498	472	446	476	485	478	532	542	514	535	567	532
5	484	483	480	473	475	482	517	539	570	530	557	585	534

Среднее арифметическое количества совпадений биграмм во всех парах 505. Прделаем ту же операцию сравнения пяти отрывков Аксакова с отрывками из произведений Апухтина и затем получим среднее арифметическое: 544. Аналогично поступим с текстами А. Бестужева (среднее арифметическое 437), Боборыкина (496). Из сравнения с четырьмя авторами тоже можно получить среднее арифметическое число совпадений биграмм, так сказать среднее средних: **496**, то есть совпадений в среднем еще примерно в полтора раза меньше, чем при сравнении произведений одного и того же автора (697).

Это собственно то, чего мы и ожидали: одно произведение в своих частях больше похоже само на себя стилистически, чем части его на части другого произведения этого же автора и еще менее похожи на части произведений других авторов, и, главное, это соот-

<sup>13</sup> Правда, в данном случае это понятие не совсем точно. Для более точного подсчета гомогенности стиля нужно было бы сравнить все отрывки всех произведений одного автора, а не только отрывки одного произведения со всеми остальными. Так что то, что мы определили, это гомогенность стиля автора в первом приближении: все отрывки автора участвуют в сравнении, но лишь с одним произведением. Однако для наших целей этого приближения вполне достаточно.



ношение четко отражено в числах (соответственно в среднем округленно **1252**, **697** и **496**) общих биграмм в двух отрывках.

Повторим операцию, сделанную с *Записками ружейного охотника* Аксакова, с текстами некоторых других авторов.

Возьмем, например, роман А.В. Амфитеатрова *Княжна*. Сравним его части между собой:

	1	2	3	4
1		869	859	726
2	869		1015	854
3	859	1015		935
4	726	854	935	

Средний уровень совпадения биграмм **876**. То есть гомогенность *Княжны* радикально более низкая, чем гомогенность *Записок ружейного охотника* С.Т. Аксакова. Далее приводить первичные таблицы не будем, дадим сразу средние показатели. Уровень гомогенности произведений Амфитеатрова **766**. Это, конечно, ниже, чем внутри одного произведения, но снижение сравнительно небольшое. Далее посчитаем среднее арифметическое число совпадений биграмм четырех отрывков Амфитеатрова с отрывками произведений Апухтина 733, Бестужева 560, Боборыкина 723. Среднее средних: **672**.

В сравнении с другими авторами понижающий перепад и здесь небольшой, но, во-первых, он есть, а во-вторых, этот, самый маленький перепад, представляет собой, как мы увидим ниже, – крайний случай, скорее всего, связанный с общим низким уровнем гомогенности анализируемого произведения Амфитеатрова (*Княжны*). Его же *Жар-цвет* дает уже более типичную картину: внутренняя гомогенность произведения **915**, гомогенность автора **815**. Количество общих биграмм с Апухтиным в среднем 800, с Бестужевым: 594, с Боборыкиным: 762. Среднее средних: **719**.

Теперь исходная лемма поддерживается не только нашими ожиданиями, но и просчитанными примерами: **наличие большего количества общих биграмм свидетельствует о большей стилистической близости** текстов. Аналогичные сравнения по большому корпусу текстов это окончательно подтверждают, результаты

этих подсчетов доступны, их в любой момент можно проверить и перепроверить (см. таблицы сравнений<sup>14</sup>).

Всего к сравнению в первом корпусе художественной литературы<sup>15</sup> были привлечены 73 произведения русской классики XIX века (порой с небольшим перехлестом в начало XX века)<sup>16</sup>.

Все это относительные результаты и величины, призванные подтвердить зависимость уровня схожести стиля от количества совпадающих биграмм, но есть и абсолютные показатели. Например, наибольшую похожесть среди отрывков разных авторов демонстрируют повесть Л.Н. Толстого *Семейное счастье* и отрывок №7 *Униженных и оскорбленных* Ф.М. Достоевского – 1194 общих биграммы. Тот же отрывок №7 *Униженных и оскорбленных* Достоевского почти так же похож на второй отрывок из романа *Рудин* И.С. Тургенева (1192 общих биграммы); то же *Семейное счастье* Толстого и отрывок №2 *Униженных и оскорбленных* Достоевского имеют 1185 одинаковых биграмм. Ниже приводим таблицу самых похожих отрывков:

Толстой. <i>Семейное счастье</i>	Достоевский. <i>Униженные и оскорбленные_7</i>	1194
Достоевский. <i>Униженные и оскорбленные_7</i>	Тургенев. <i>Рудин_2</i>	1192
Толстой. <i>Семейное счастье</i>	Достоевский. <i>Униженные и оскорбленные_2</i>	1185
Достоевский. <i>Униженные и оскорбленные_7</i>	Тургенев. <i>Накануне_2</i>	1180
Салтыков-Щедрин. <i>Повести_6</i>	Достоевский. <i>Униженные и оскорбленные_5</i>	1180
Тургенев. <i>Накануне_2</i>	Достоевский. <i>Униженные и оскорбленные_7</i>	1180

<sup>14</sup> <https://github.com/nevmenandr/ngrams-compare-style/blob/master/fiction/15000/tables/partition2gram.tsv> Б.В. Орехов, И.В. Пешков, fiction results, (19.02.2021)

<sup>15</sup> Как минимум два из них имеют не до конца художественный характер (*История государства российского* Карамзина и *История пугачевского бунта* Пушкина), что нужно иметь в виду, однако принципиально на общей картине это не сказывается.

<sup>16</sup> <https://cloud.mail.ru/public/F857/k5M8WooFy> Б.В. Орехов, И.В. Пешков, fiction, (19.02.2021)

Достоевский. <i>Униженные и оскорбленные</i> _5	Салтыков-Щедрин. <i>Повести</i> _6	1180
Достоевский. <i>Униженные и оскорбленные</i> _6	Салтыков-Щедрин. <i>Повести</i> _6	1164
Достоевский. <i>Униженные и оскорбленные</i> _4	Салтыков-Щедрин. <i>Повести</i> _6	1163
Толстой. <i>Семейное счастье</i>	Достоевский. <i>Униженные и оскорбленные</i> _6	1160
Достоевский. <i>Униженные и оскорбленные</i> _7	Толстой. <i>Семейное счастье</i> _1	1159
Достоевский. <i>Униженные и оскорбленные</i> _6	Тургенев. <i>Рудин</i> _2	1156
Достоевский. <i>Униженные и оскорбленные</i> _4	Толстой. <i>Юность</i> _2	1154
Достоевский. <i>Униженные и оскорбленные</i> _7	Салтыков-Щедрин. <i>Повести</i> _6	1154
Толстой. <i>Семейное счастье</i>	Достоевский. <i>Униженные и оскорбленные</i> _5	1152
Тургенев. <i>Рудин</i> _2	Достоевский. <i>Идиот</i> _13	1151
Толстой. <i>Семейное счастье</i>	Достоевский. <i>Униженные и оскорбленные</i> _4	1150

Сейчас для нас существенна не сама по себе высшая похожесть, а то, что число **1194** является крайней отметкой возможной общности отрывков разных авторов: более высокие показатели имеют пары текстов исключительно одного и того же авторства.

Толстой. <i>Юность</i> _2	Толстой. <i>Детство</i> _2	1325
Толстой. <i>Т.2. Война и мир</i> _7	Толстой. <i>Т.1 Война и мир</i> _6	1325
Достоевский. <i>Игрок</i> _3	Достоевский. <i>Униженные и оскорбленные</i> _7	1317
Достоевский. <i>Идиот</i> _2	Достоевский. <i>Униженные и оскорбленные</i> _6	1317
Толстой. <i>Т.1. Война и мир</i> _6	Толстой. <i>Т.2. Война и мир</i> _5	1311
Достоевский. <i>Униженные и оскорбленные</i> _4	Достоевский. <i>Игрок</i> _3	1306
Толстой. <i>Анна Каренина</i> _2	Толстой. <i>Т.2. Война и мир</i> _5	1302
Толстой. <i>Т.2. Война и мир</i> _7	Толстой. <i>Т.4. Война и мир</i> _6	1300

И так далее: уровень в 1194 общих биграмм превышает результаты сравнения 164-х пар отрывков произведений одного и того же автора. Все это позволяет сделать предположение, что в принципе существует ощутимая граница, выше которой общность разных авторов не идет. Для наших 450 отрывков художественных произведений первого корпуса это уровень в 1194 биграммы. Так и сформулируем: 1194 одинаковых биграмм – **верхний количественный предел присутствия одинаковых биграмм для разных авторских стилей** или, сокращенно, **верхняя стилистическая граница разных авторов**.

Есть еще и нижний уровень индивидуального стиля:

Карамзин. <i>История государства Российского_1</i>	Карамзин. <i>Повести_1</i>	485
Бестужев. <i>Фрегат Надежда_1</i>	Бестужев. <i>Роман и Ольга_1</i>	462
Бестужев. <i>Лейтенант Белозор_1</i>	Бестужев. <i>Роман и Ольга_1</i>	444
Бестужев. <i>Роман и Ольга_1</i>	Бестужев. <i>Лейтенант Белозор_1</i>	421
Бестужев. <i>Роман и Ольга_1</i>	Бестужев. <i>Ревельский турнир_1</i>	386

Меньше 386 общих биграмм в отрывках одного и того же автора не наблюдалось. Значит, рискнем сделать обобщение, **меньшее количество совпадений с высокой степенью вероятности говорит о том, что авторы разные**.

Разумеется, полученные статистические данные требуют проверки на большем массиве текстов, и частично мы проделали такую проверку, сравнив отрывки еще одного корпуса художественной литературы<sup>17</sup>, на сей раз не вполне хрестоматийной классики.

Вот первая десятка пар разных авторов:

Степняк-Кравчинский. <i>Подпольная Россия_3</i>	Якубович. Т.2. <i>В мире отверженных_5</i>	1050
Степняк-Кравчинский. <i>Подпольная Россия_3</i>	Кущевский. <i>Негорев_6</i>	1029

<sup>17</sup> <https://cloud.mail.ru/public/Eiae/ZPLQvLLje>\_Б.В. Орехов, И.В. Пешков, bakhtin\_authorship\_fiction\_corpus, (19.02.2021)

Якубович. Т.2. <i>В мире отверженных_5</i>	Решетников. <i>Между людьми_4</i>	1011
Якубович. Т.2. <i>В мире отверженных_5</i>	Решетников. <i>Между людьми_1</i>	993
Решетников. <i>Между людьми_4</i>	Куцевский. <i>Негорев_6</i>	993
Решетников. <i>Между людьми_4</i>	Куцевский. <i>Негорев_5</i>	988
Якубович. Т.2. <i>В мире отверженных_5</i>	Куцевский. <i>Негорев_6</i>	985
Соловьев. <i>Жених царевны_3</i>	Степняк-Кравчинский. <i>Штундист_3</i>	984
Решетников. <i>Между людьми_2</i>	Якубович. Т.2. <i>В мире отверженных_5</i>	971
Решетников. <i>Между людьми_4</i>	Степняк-Кравчинский. <i>Подпольная Россия_3</i>	968

Тут крайний уровень допустимой похожести разных авторов существенно ниже, чем у хрестоматийной классики, всего **1050** (против 1194 в первом корпусе) и в первой же десятке падает до 968, в то время как в классическом корпусе он и через 20 пар падает меньше (до 1136), что в совокупности свидетельствует о количественно большем уровне гомогенности, так сказать, высокой классики по сравнению с классикой, не столь общепризнанно высокой.

Нижний уровень второго корпуса (607), наоборот, существенно выше первого (386).

Кржижановский. <i>Возвращение Мюнхгаузена_2</i>	Кржижановский. <i>Сальвер-Гюль_1</i>	607
---	--------------------------------------	-----

Соответственно диапазон *смешанной* зоны для нехрестоматийной классики уменьшается, что не отменяет наличия абсолютных и относительных границ индивидуального стиля: с показателями выше смешанной зоны возможен один авторский стиль в сравниваемых текстах, ниже – не менее двух.

Обратившись к нашей частной задаче (аутентификации спорных текстов Бахтина), мы проделали вышеописанные сравнительные

операции уже для корпуса научных текстов<sup>18</sup>. Тут проблема слегка осложнялась взаимной цитатностью у некоторых авторов. Так, цитирование одного и того же абзаца из *Дневника писателя* Ф.М. Достоевского в *МФЯ* и *Мышлении и речи* Л.С. Выготского привело к резкому увеличению общих биграмм при их сравнении<sup>19</sup>. В этих случаях приходилось вводить для анализа те же произведения без цитат и соответственно только их и учитывать в выводах. Это коснулось прежде всего работ Л.С. Выготского и А.А. Потебни, перегруженных цитатами<sup>20</sup>. Кроме того, и ключевые для нашего анализа тексты (*МФЯ*, *ФМЛ* и *Фрейдизм*) были анализированы дополнительно в вариантах, очищенных от цитат.

Сразу же стоит подчеркнуть, что количественный порог общих биграмм, который не переступают отрывки разных авторов, для научно-гуманитарных текстов значительно ниже, чем для таких же по размеру (напоминаем, что речь все время идет об отрывках в 15000 слов) классических текстов художественной литературы. Если для последних этот порог составляют **1194** биграммы, то для первых (и для нашего корпуса точно) он не превышает **880** биграмм. То есть научные тексты в целом стилистически менее гомогенны. Это, конечно, можно теоретически объяснить, но сейчас это не входит в нашу задачу. Наша задача сейчас понять, почему последний порог допустимости разного авторства радикально переступают нижеперечисленные пары:

Волошинов. <i>МФЯ</i> без цитат_2	Медведев. <i>ФМЛ</i> без цитат_3	1037
Волошинов. <i>МФЯ</i> без цитат_2	Медведев. <i>ФМЛ</i> _3	1036
Волошинов. <i>МФЯ</i> без цитат_2	Медведев. <i>ФМЛ</i> без цитат_2	963
Волошинов. <i>МФЯ</i> _2	Медведев. <i>ФМЛ</i> без цитат_2	955
Бахтин. <i>Проблема содержания,</i>	Медведев. <i>ФМЛ</i> без	945

<sup>18</sup> <https://cloud.mail.ru/public/9pA4/7pqjbpFeX>\_Б.В. Орехов, И.В. Пешков, corpus, (19.02.2021)

<sup>19</sup> Указанная цитата из Достоевского дала, например, 674 биграммы, некоторые из них, могли повториться, но вряд ли эти повторы существенно уменьшили их количество.

<sup>20</sup> Вариант без цитат проанализирован также для работы *Проблема символа и реалистическое искусство* А.Ф. Лосева.

<i>материала и формы_1</i>	цитат_3	
Бахтин. <i>Проблема содержания, материала и формы_1</i>	Медведев. <i>ФМЛ_3</i>	944
Волошинов. <i>МФЯ без цитат_2</i>	Медведев. <i>ФМЛ без цитат_1</i>	944
Бахтин. <i>Проблемы творчества Достоевского_1</i>	Медведев. <i>ФМЛ без цитат_3</i>	942
Волошинов. <i>МФЯ без цитат_2</i>	Медведев. <i>ФМЛ_2</i>	941
Волошинов. <i>МФЯ без цитат_2</i>	Медведев. <i>ФМЛ_1</i>	941
Бахтин. <i>Проблема содержания, материала и формы_1</i>	Медведев. <i>ФМЛ_1</i>	937
Волошинов. <i>МФЯ_2</i>	Медведев. <i>ФМЛ_2</i>	934
Волошинов. <i>МФЯ_2</i>	Медведев. <i>ФМЛ без цитат_1</i>	934
Волошинов. <i>МФЯ_2</i>	Медведев. <i>ФМЛ_1</i>	932
Бахтин. <i>Проблема содержания, материала и формы_1</i>	Медведев. <i>ФМЛ без цитат_1</i>	931

Причем и граница в 880 биграмм, скорее всего, завышенная. Приводим пограничные пары.

Лосев. <i>Проблема символа без цитат_2</i>	Потебня. <i>Мысль и язык без цитат_2</i>	880
Лосев. <i>Проблема символа без цитат_2</i>	Потебня. <i>Мысль и язык без цитат_3</i>	875
Лосев. <i>Проблема символа без цитат_2</i>	Медведев. <i>ФМЛ_2</i>	871
Потебня. <i>Мысль и язык_3</i>	Лосев. <i>Проблема символа_1</i>	870
Лосев. <i>Проблема символа_2</i>	Медведев. <i>ФМЛ_2</i>	869

Возможно, высокий уровень похожести тут определяется особенностями письма слепого А.Ф. Лосева. В таком случае, можно предполагать, что черта индивидуального стиля пролегает гораздо ниже, примерно на отметке **868** биграмм, когда впервые, если не считать текстов «Бахтина под маской», в парах сравнения появляется не Лосев, а другие авторы<sup>21</sup>.

<sup>21</sup> Мы исходим из того, что все точно небахтинские тексты со взаимными (или одинаковыми) цитатами выводятся из анализа.

Выготский. <i>Мышление и речь</i> без цитат_7	Потебня. <i>Мысль и язык</i> _1	868
Потебня. <i>Мысль и язык</i> без цитат_1	Выготский. <i>Мышление и речь</i> без цитат_7	865

Поэтому продолжим предъявление всех (!) отрывков, превышающих этот, теневой рубеж единства стиля (868):

Волошинов. <i>МФЯ</i> без цитат_2	Бахтин. <i>Слово в романе</i> _1	923
Волошинов. <i>МФЯ</i> _1	Медведев. <i>ФМЛ</i> без цитат_2	916
Волошинов. <i>МФЯ</i> без цитат_1	Медведев. <i>ФМЛ</i> без цитат_2	916
Волошинов. <i>МФЯ</i> _1	Медведев. <i>ФМЛ</i> без цитат_2	916
Волошинов. <i>МФЯ</i> _2	Бахтин. <i>Слово в романе</i> _1	911
Волошинов. <i>МФЯ</i> _1	Медведев. <i>ФМЛ</i> _2	903
Волошинов. <i>МФЯ</i> без цитат_1	Медведев. <i>ФМЛ</i> _2	903
Бахтин. <i>Проблемы поэтики Достоевского</i> _1	Медведев. <i>ФМЛ</i> без цитат_3	889
Бахтин. <i>Проблема содержания, материала и формы</i> _1	Медведев. <i>ФМЛ</i> без цитат_2	889
Бахтин. <i>Проблемы творчества Достоевского</i> _1	Медведев. <i>ФМЛ</i> _3	884
Волошинов. <i>МФЯ</i> без цитат_1	Волошинов. <i>Стилистика художественной речи</i> _1	877
Бахтин. <i>Слово в романе</i> _1	Медведев. <i>ФМЛ</i> без цитат_2	875
Бахтин. <i>Проблемы творчества Достоевского</i> _2	Медведев. <i>ФМЛ</i> без цитат_3	873
Волошинов. <i>МФЯ</i> без цитат_2	Бахтин. <i>Проблемы творчества Достоевского</i> _2	871
Бахтин. <i>Проблема содержания, материала и формы</i> _1	Медведев. <i>ФМЛ</i> _2	871
Волошинов. <i>МФЯ</i> без цитат_2	Бахтин. <i>Проблемы творчества Достоевского</i> _2	871

Итак, еще 16 сравниваемых пар текстов (из которых 6 выше лишь теневого рубежа), титульно атрибутированных Медведеву, Во-



лошинову и Бахтину, оказываются с высокой степенью вероятности написанными одной и той же рукой.

И прямо рядом с этой границей (868) имеем еще 6 пар в той же композиции:

Волошинов. <i>МФЯ</i> без цитат_2	Бахтин. <i>Слово в романе</i> _2	867
Бахтин. <i>Проблемы творчества Достоевского</i> _2	Медведев. <i>ФМЛ</i> _3	867
Бахтин. <i>Проблемы творчества Достоевского</i> _1	Медведев. <i>ФМЛ</i> _1	865
Бахтин. <i>Проблемы творчества Достоевского</i> _1	Медведев. <i>ФМЛ</i> без цитат_1	864
Волошинов. <i>МФЯ</i> _2	Бахтин. <i>Проблемы творчества Достоевского</i> _2	860
Бахтин. <i>Слово в романе</i> _1	Медведев. <i>ФМЛ</i> без цитат_3	860

Объяснения всех этих аномалий напрашивается само собой.

Однако обратимся ко второму корпусу научно-гуманитарных текстов<sup>22</sup> (хотя следует заметить, что здесь привлекались и некоторые переходные жанры, в частности биографические, философские, которые как раз фигурируют в этой таблице), сравнение отрывков из которых несколько повышает нехудожественный стилевой порог:

Чаадаев. <i>Отрывки</i> _1	Страхов. <i>Мир как целое</i> _7	969
Шестов. <i>Добро</i> _2	Вересаев. <i>Да здравствует</i> _3	965
Шестов. <i>Добро</i> _1	Вересаев. <i>Да здравствует</i> _3	954
Александров. <i>Байрон</i> _1	Каменский. <i>Авраам Линкольн</i> _1	941
Александров. <i>Байрон</i> _2	Каменский. <i>Авраам Линкольн</i> _1	941
Страхов. <i>Мир как целое</i> _4	Чаадаев. <i>Отрывки</i> _1	940
Чаадаев. <i>Отрывки</i> _1	Страхов. <i>Мир как целое</i> _2	940

<sup>22</sup> <https://cloud.mail.ru/public/6LAJ/HNEmGhS6v> Б.В. Орехов, И.В. Пешков, bakhtin\_authorship\_nonfiction\_corpus (19.02.2021)

Результаты сравнения Вересаева и Шестова можно смело не учитывать, так как их тексты посвящены Л.Н. Толстому и имеют общие цитаты из последнего. Остаются пары Страхов–Чаадаев и Александров–Каменский соответственно с философскими и биографическими текстами. Гуманитарно-научными их можно назвать лишь условно.

Но для чистоты эксперимента согласимся с имеющимися порогами: **880** по первому корпусу нехудожественных текстов и **969** – по второму. Даже последний, заведомо завышенный рубеж преодолевают 8 пар:

Волошинов. <i>МФЯ</i> _1	Медведев. <i>ФМЛ</i> без цитат_1	<b>1034</b>
Волошинов. <i>МФЯ</i> без цитат_1	Медведев. <i>ФМЛ</i> _1	<b>1031</b>
Волошинов. <i>МФЯ</i> _2	Медведев. <i>ФМЛ</i> без цитат_3	<b>1031</b>
Волошинов. <i>МФЯ</i> _2	Медведев. <i>ФМЛ</i> _3	<b>1023</b>
Волошинов. <i>МФЯ</i> без цитат_1	Медведев. <i>ФМЛ</i> без цитат_3	<b>980</b>
Волошинов. <i>МФЯ</i> _1	Медведев. <i>ФМЛ</i> без цитат_3	<b>980</b>
Волошинов. <i>МФЯ</i> _1	Медведев. <i>ФМЛ</i> _3	<b>975</b>
Волошинов. <i>МФЯ</i> без цитат_1	Медведев. <i>ФМЛ</i> _3	<b>975</b>

Это преодоление снимает всяческие сомнения в единстве авторства всех отрывков *МФЯ* и *ФМЛ*. Небольшие сомнения остаются лишь в связи со второй частью *ФМЛ*, не преодолевшей завышенный стилиевой порог, но от этого порога не слишком отдалившейся:

Волошинов. <i>МФЯ</i> _1	Медведев. <i>ФМЛ</i> без цитат_2	916
Волошинов. <i>МФЯ</i> без цитат_1	Медведев. <i>ФМЛ</i> без цитат_2	916
Волошинов. <i>МФЯ</i> _1	Медведев. <i>ФМЛ</i> без цитат_2	916

Таким образом, обе части *МФЯ* и две из трех частей *ФМЛ* в сравнении друг с другом преодолевают порог числа совпадающих биграмм (969), выше которого не бывает разных авторов, вторая часть *ФМЛ* находится рядом с порогом. Самую высокую стилистическую похожесть демонстрируют первые части обеих книг (1034) и третья часть *ФМЛ* со второй частью *МФЯ* (1031). Вывод с высочайшей степенью вероятности: *МФЯ* и *ФМЛ* написаны одним и тем же человеком.

Определить, что это за человек, не так трудно: его фамилия тоже появляется в запороговых и околопороговых парах. На *Проблему содержания, материала и формы в словесном художественном творчестве*, текст Бахтина середины 20-х годов, который не был опубликован до 1975 года, недопустимо для разных авторов первого корпуса похожа третья часть ФМЛ (945) и первая часть ФМЛ (937). Кроме того, примерно такую же схожесть демонстрируют первая часть *Проблем творчества Достоевского* и третья часть ФМЛ (942).

Медведев. ФМЛ без цитат_3	Бахтин. <i>Проблема содержания, материала и формы_1</i>	945
Медведев. ФМЛ без цитат_3	Бахтин. <i>Проблемы творчества Достоевского_1</i>	942
Медведев. ФМЛ_2	Волошинов. МФЯ без цитат_2	941
Медведев. ФМЛ_1	Волошинов. МФЯ без цитат_2	941
Медведев. ФМЛ_1	Бахтин. <i>Проблема содержания, материала и формы_1</i>	937
Медведев. ФМЛ_3	Бахтин. <i>Проблемы творчества Достоевского_1</i>	884

Другие пары, включающие фигурантов дела «Бахтин под маской», показывают следующие результаты:

Бахтин. <i>Проблема содержания, материала и формы_1</i>	Медведев. ФМЛ_2	871
Бахтин. <i>Проблемы творчества Достоевского_2</i>	Медведев. ФМЛ_3	867
Бахтин. <i>Проблемы творчества Достоевского_1</i>	Медведев. ФМЛ_1	865
Бахтин. <i>Проблемы творчества Достоевского_2</i>	Волошинов. МФЯ_2	860
Бахтин. <i>Слово в романе_1</i>	Медведев. ФМЛ_3	858
Бахтин. <i>Слово в романе_1</i>	Медведев. ФМЛ_2	855
Бахтин. <i>Проблемы поэтики Достоевского_1</i>	Медведев. ФМЛ_3	853
Бахтин. <i>Проблема содержания, материала и формы</i>	Медведев. ФМЛ_2	848
Волошинов. <i>Стилистика художественной речи</i>	Медведев. ФМЛ_3	847

Бахтин. Слово в романе_2	Волошинов. МФЯ_2	846
Бахтин. Слово в романе_2	Волошинов. МФЯ_2	846
Бахтин. Слово в романе_2	Медведев. ФМЛ_3	840
Бахтин. Проблема содержания, материала и формы_1	Волошинов. МФЯ_2	839

Все эти 13 пар находятся в смешанной зоне (в которую попадают тексты как одинаковых, так и разных авторов), но близки к порогу единого авторства (880 для первого корпуса и 969 для второго), а также – рядом с теневым, более адекватным, порогом первого корпуса (865) или выше него. В 12 парах имя Бахтин фигурирует рядом с именами Волошинова и Медведева. Существенно, что последняя строка показывает всего лишь 207-ю пару сверху по уровню схожести из 8129 сравненных пар первого корпуса нонфикшн (см. таблицы<sup>23</sup>).

Любопытно также, что «Волошинов» порой меньше похож сам на себя, чем на «Медведева», хотя тоже похож: в разных сочетаниях отрывков четыре пары «Фрейдизма» и МФЯ показывают количество общих биграмм 812, 830, 850 и 861. Последние два результата совсем близки к теневому порогу первого корпуса.

При этом единственная большая несомненно медведевская работа *В лаборатории писателя* показывает в сравнении с ФМЛ числа, далекие от порога единства стиля (для надежности мы даем эту работу в двух вариантах: последней строкой приводится та же самая работа, но без цитат):

	ФМЛ1	ФМЛ2	ФМЛ3
<i>Творческая деятельность1</i>	685	712	687
<i>Творческая деятельность2</i>	668	701	711
<i>В лаборатории писателя</i>	685	733	724

Еще более выразительные результаты получены по сравнению нехудожественных отрывков в 20000 слов.

Волошинов. МФЯ без цитат_1	Медведев. ФМЛ без цитат_2	1404	
Волошинов. МФЯ_1	Медведев. ФМЛ без цитат_2	1404	

<sup>23</sup> <https://github.com/nevmenandr/ngrams-compare-style/tree/master/take12.2018/nonfiction2/15000/tables> Б.В. Орехов, И.В. Пешков, ngrams-compare-style (19.02.2021).

Волошинов. МФЯ_1	Медведев. ФМЛ_2	1387	
Волошинов. МФЯ без цитат_1	Медведев. ФМЛ_2	1387	
Медведев. ФМЛ без цитат_2	Волошинов. МФЯ без цитат_2	1380	
Волошинов. МФЯ_1	Медведев. ФМЛ_1	1377	
Волошинов. МФЯ без цитат_1	Медведев. ФМЛ_1	1377	
Волошинов. МФЯ_2	Медведев. ФМЛ без цитат_2	1368	
Волошинов. МФЯ без цитат_2	Медведев. ФМЛ_2	1364	
Волошинов. МФЯ_1	Медведев. ФМЛ без цитат_1	1361	
Волошинов. МФЯ без цитат_1	Медведев. ФМЛ без цитат_1	1361	
Волошинов. МФЯ_2	Медведев. ФМЛ_2	1352	
Волошинов. МФЯ без цитат_2	Бахтин. Проблемы творчества Достоевского_2	1272	<b>1337</b> nF2
Волошинов. МФЯ_2	Бахтин. Проблемы творчества Достоевского_2	1268	
Медведев. ФМЛ без цитат_2	Бахтин. Слово в романе_1	1261	
Волошинов. МФЯ без цитат_2	Бахтин. Слово в романе_1	1258	
Медведев. ФМЛ без цитат_2	Бахтин. Проблемы творчества Достоевского_1	1243	
Волошинов. МФЯ_2	Бахтин. Слово в романе_1	1237	
Медведев. ФМЛ_2	Бахтин. Слово в романе_1	1234	
Медведев. ФМЛ_1	Бахтин. Проблемы творчества Достоевского_1	1204	
Лосев. Проблема символа_1	Потебня. Мысль и язык без цитат_2	<b>1201</b>	

Абсолютный стилевой порог в 1337 биграмм, который был получен по второму корпусу нонфикшн (по первому, порог составляет

1201 биграмма) преодолевают во взаимном сравнении обе части *МФЯ* и обе части *ФМЛ*, то есть относительно всего текста *ФМЛ* снимаются последние сомнения. Между первым и вторым порогом расположились тексты двух работ Бахтина в сравнении с *МФЯ* и *ФМЛ*. На всякий случай напомним, что статья Бахтина *Проблема содержания, материала и формы в словесном художественном творчестве* имеет объем меньше 20000 слов и в этих сравнениях участия не принимает.

Бахтинское присутствие в текстах *Бахтина под маской* мы проверили еще дополнительным сравнением полного словника биграмм всех произведений Бахтина с двумя корпусами нехудожественных текстов. Приведем верхнюю часть таблиц результатов по 15000-м отрывкам. По первому корпусу:

Медведев. <i>ФМЛ</i> без цитат_3	3549
Медведев. <i>ФМЛ</i> _3	3453
Медведев. <i>ФМЛ</i> _1	3411
Волошинов. <i>МФЯ</i> без цитат_2	3398
Медведев. <i>ФМЛ</i> без цитат_1	3391
Волошинов. <i>МФЯ</i> _2	3385
Медведев. <i>ФМЛ</i> без цитат_2	3312
Медведев. <i>ФМЛ</i> _2	3310
Волошинов. <i>МФЯ</i> без цитат_1	3236
Волошинов. <i>МФЯ</i> _1	3235
<b>Выготский. Психология искусства без цитат_5</b>	<b>3091</b>
Выготский. <i>Психология искусства</i> _1	3068
Выготский. <i>Психология искусства</i> _6	3050
Выготский. <i>Психология искусства</i> _2	3047
Волошинов. <i>Стилистика художественной речи</i> _1	3041
Выготский. <i>Психология искусства</i> без цитат_1	3013
Выготский. <i>Психология искусства</i> без цитат_2	3010
Медведев. <i>Творческая деятельность</i> _1	2998
Выготский. <i>Психология искусства</i> без цитат_4	2985
Лосев. <i>Проблема символа</i> _4	2972

По второму корпусу:

Чаадаев. <i>Отрывки</i> _1	3040
Страхов. <i>Мир как целое</i> _3	2913
Страхов. <i>Мир как целое</i> _1	2858

Страхов. <i>Мир как целое</i> _5	2847
Шестов. <i>Добро</i> _1	2818
Страхов. <i>Мир как целое</i> _2	2818
Страхов. <i>Мир как целое</i> _7	2818
Шестов. <i>Добро</i> _2	2810
Фрейденберг. <i>Семантика сюжета и жанра</i> _1	2780
Булгаков. <i>Свет невечерний</i> _3	2738
Пропп. <i>Фольклор. Литература</i> _4	2726

Тут тоже проявилась граница, на этот раз четко отделяющая точно бахтинские тексты от небахтинских или не обязательно бахтинских. Существенно выше этой границы оказались все отрывки *МФЯ* и *ФМЛ*, что дает еще одно подтверждение того, что они полностью написаны Бахтиным.

Итак, на основании проделанного анализа с максимально высокой степенью вероятности установлено:

1. Автор *ФМЛ* и автор *МФЯ* – один и тот же человек.
2. Автор третьей части *ФМЛ* и атрибутированных работ Бахтина – один и тот же человек.
3. Автор второй части *МФЯ* и атрибутированных работ Бахтина – один и тот же человек.

С высочайшей степенью вероятности можно признать, что:

4. Автор первой части *ФМЛ* и атрибутированных работ Бахтина – один и тот же человек.

И с высокой степенью вероятности признаем:

5. Автор *ФМЛ* и П.Н.Медведев – разные люди.
6. Автор второй части *ФМЛ* и атрибутированных работ Бахтина – один и тот же человек.

Вывод из этих посылок в целом имеет максимальную вероятность: **М.М. Бахтин единоличный автор книг: *МФЯ* (*Марксизм и философия языка*) и *ФМЛ* (*Формальный метод в литературоведении*).**

Из-за некоторой тематической обособленности чуть сложнее обстоит дело с атрибуцией *Фрейдизма*, хотя единый автор с *МФЯ* доказывается не только титульным единством маски «Волошинов». Выше мы продемонстрировали, что *Фрейдизм* и обе части *МФЯ* имеют хорошие предпороговые цифры общности.

К аналогичным выводам приводит и Delta. В корпусе нехудожественных текстов числовые значения дистанции между *МФЯ*

и *ФМЛ* 0,469 (чем меньше число, тем ближе тексты стилистически), что практически однозначно свидетельствует о том, что тексты принадлежат одному автору. Ср. значение 0,68 для двух книг В. Проппа о сказке. Иными словами, книги фольклориста похожи друг на друга даже несколько меньше, чем *МФЯ* и *ФМЛ*. Между тем, дистанция от *ФМЛ* до текстов, несомненно написанных П.Н. Медведевым, колеблется в пределах от 0,9 до 1,5. Результаты *Delta* трактуются однозначно: Медведев не может быть автором *ФМЛ*. Расстояние от *МФЯ* до ранних работ Бахтина примерно 0,7, то есть такое же, как между разными книгами одного автора. Следовательно, именно Бахтина следует считать автором работ «под маской»<sup>24</sup>.

Добавим еще, возвращаясь к результатам исследований по нашей методике, что для научных, как и для художественных, текстов выделяется смешанная зона результатов, когда количество одинаковых биграмм у сравниваемых отрывков не дает оснований однозначно судить, принадлежит ли текст одному и тому же или разным авторам. Эта смешанная зона имеет не только верхнюю, но и нижнюю границу. Самым стилистически разнообразным автором, у которого в отрывках разных работ присутствует лишь 458 общих биграмм, оказывается М.М. Бахтин, причем с большим отрывом от других авторов первого корпуса:

Фрейденберг. <i>Семантика сюжета и жанра_4</i>	Фрейденберг. <i>Поэтика сюжета и жанра_1</i>	552
Бахтин. <i>Творчество Франсуа Рабле_4</i>	Бахтин. <i>Слово в романе_2</i>	549
Бахтин. <i>Творчество Франсуа Рабле_3</i>	Бахтин. <i>К философии поступка_1</i>	547
Бахтин. <i>Формы времени и хромотона_3</i>	Бахтин. <i>К философии поступка_1</i>	546
Бахтин. <i>Творчество Франсуа Рабле_1</i>	Бахтин. <i>К философии поступка_1</i>	545
Тюпа. <i>Литература и ментальность_2</i>	Тюпа. <i>Дискурс и жанр_1</i>	542
Бахтин. <i>Творчество Франсуа Рабле_5</i>	Бахтин. <i>К философии поступка_1</i>	527

<sup>24</sup> Мы приносим благодарность Б.В. Орехову за осмысление результатов анализа по Дельте, а также за помощь в подсчетах по всей статье.



Бахтин. <i>Творчество Франсуа Рабле_2</i>	Бахтин. <i>К философии поступка_1</i>	526
Бахтин. <i>Творчество Франсуа Рабле_6</i>	Бахтин. <i>К философии поступка_1</i>	524
Бахтин. <i>Проблемы поэтики Достоевского_3</i>	Бахтин. <i>К философии поступка_1</i>	516
Бахтин. <i>Творчество Франсуа Рабле_8</i>	Бахтин. <i>К философии поступка_1</i>	513
Бахтин. <i>Творчество Франсуа Рабле_7</i>	Бахтин. <i>К философии поступка_1</i>	512
Бахтин. <i>Творчество Франсуа Рабле_4</i>	Бахтин. <i>Проблема содержания, материала и формы_1</i>	512
Бахтин. <i>Творчество Франсуа Рабле_4</i>	Бахтин. <i>К философии поступка_1</i>	458

Во втором корпусе по непохожести собственных текстов доминирует О.М. Фрейденоберг (обратим внимание, что ее тексты присутствуют в обоих корпусах):

Ткачев. <i>Немецкие идеалисты_1</i>	Ткачев. <i>Кладези_1</i>	733
Фрейденоберг. <i>Поэтика сюжета и жанра_6</i>	Фрейденоберг. <i>Семантика сюжета и жанра_1</i>	729
Фрейденоберг. <i>Поэтика сюжета и жанра_2</i>	Фрейденоберг. <i>Семантика сюжета и жанра_5</i>	633
Фрейденоберг. <i>Поэтика сюжета и жанра_3</i>	Фрейденоберг. <i>Семантика сюжета и жанра_5</i>	630
Фрейденоберг. <i>Поэтика сюжета и жанра_1</i>	Фрейденоберг. <i>Семантика сюжета и жанра_5</i>	557
Фрейденоберг. <i>Поэтика сюжета и жанра_1</i>	Фрейденоберг. <i>Семантика сюжета и жанра_4</i>	552

В целом ниже (то есть в смысловом отношении выше) бахтинских показателей стилистического разнообразия (458) единства стиля уже не наблюдается. Причем это касается как нехудожественных, так и «неклассических» художественных текстов. На нашем материале это почти абсолютная нижняя граница (цифры ниже присутствуют

только для трех пар отрывков произведений А.А. Бестужева-Марлинского<sup>25</sup>).

Тот факт, что Бахтин как автор был способен на максимальное стилистическое разнообразие, придает еще большую значимость превышению текстами *Бахтина под маской* верхней границы смешанной зоны, показатели выше которой с очень большой степенью вероятности гарантируют единство стиля сравниваемых отрывков. М.М. Бахтин умел писать разнообразно, но создавая общую концепцию языка и речевого произведения, думал именно об этой концепции, и не старался стилистически завуалировать свое авторство, надевая внешнюю идеологическую маску, зафиксированную фамилиями его друзей. Или при всем старании завуалировать свою индивидуальность в стиле оказалось невозможно.

## REFERENCES:

- Burrows John: *'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship*, „Literary and Linguistic Computing“ 1 September 2002, Volume 17, Issue 3, p.267-287, <https://doi.org/10.1093/lc/17.3.267>
- Eisen Mark, Ribeiro Alejandro, Segarra Santiago, Egan Gabriel: *Stylometric analysis of Early Modern period English plays*, „Digital Scholarship in the Humanities“ 2018, v. 33, issue 3, 1 September, p.500-528, <https://doi.org/10.1093/lc/fqx059>
- Gladwin Alexander, Lavin Matthew, Look Daniel: *Stylometry and collaborative authorship: Eddy, Lovecraft, and 'The Loved Dead'*, „Digital Scholarship in the Humanities“, 2017, v. 32, issue 1, 1 April, p. 123-140, <https://doi.org/10.1093/lc/fqv026>
- Ilsemann Hartmut: *Stylometry approaching Parnassus*, „Digital Scholarship in the Humanities“ 2018, v. 33, issue 3, 1 September, p.548-556, <https://doi.org/10.1093/lc/fqx058>
- Khmelev Dmitriy: *Raspoznavaniye avtora teksta s ispol'zovaniyem tsepey A.A. Markova*, „Vestnik MGU“ 2000, ser. 9: Filologiya, № 2, s.115-126 (Хмелёв Дмитрий: *Распознавание автора текста с использованием цепей А.А. Маркова*, „Вестник МГУ“ 2000, сер. 9: Филология, № 2, с.115-126).

<sup>25</sup> Причем, во всех этих парах присутствует историческая повесть *Роман и Ольга*, текст которой стилизован под древнерусский.

- Kukushkina Ol'ga, Polikarpov Anatoliy, Khmelev Dmitriy: *Opredeleniye avtorstva teksta s ispol'zovaniyem bukvennoy i grammaticheskoy informatsii*, „Problemy peredachi informatsii“ 2001, t. 37, № 2, s.96-108 (Кукушкина Ольга, Поликарпов Анатолий, Хмелёв Дмитрий: *Определение авторства текста с использованием буквенной и грамматической информации*, “Проблемы передачи информации” 2001, т. 37, № 2, с. 96-108).
- Lutoslawski Wincenty: *Principes de stylométrie appliqués à la chronologie des œuvres de Platon*, „Revue des Études Grecques“, 1898, t. 11, fascicule 41, p. 61-81.
- Marusenko Mikhail: *Atributsiya anonimnykh i psevdonimnykh literaturnykh proizvedeniy metodami teorii raspoznavaniya obrazov*. Leningrad 1990 (Марусенко Михаил: *Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов*. Ленинград 1990).
- Mendenhall Thomas: *The characteristic curves of composition*, „Science“ 1887, 9 (214), p. 237-249.
- Morozov Nikolay: *Lingvisticheskiye spektry: sredstvo dlya otlicheniya plagiatov ot istinnykh proizvedeniy togo ili inogo izvestnogo avtora. Stilemetricheskii etjud*, „Izvestiya otd. russkogo yazyka i slovesnosti Imp. Akad. Nauk“ 1915, t. XX, kn. 4 (Морозов Николай: *Лингвистические спектры: средство для отличения плагиатов от истинных произведений того или иного известного автора. Стилеметрический этюд*, „Известия отд. русского языка и словесности Имп. Акад. Наук“ 1915, т. XX, кн. 4).
- Nini Andrea: *An authorship analysis of the Jack the Ripper letters*, „Digital Scholarship in the Humanities“, 2018, v. 33, issue 3, 1 September, p.621-636, <https://doi.org/10.1093/lc/fqx065>
- Oakes Michael: *Computer stylometry of C. S. Lewis's The Dark Tower and related texts*, „Digital Scholarship in the Humanities“, 2018, v. 33, issue 3, 1 September, p.637-650, <https://doi.org/10.1093/lc/fqx043>
- Peshkov Igor': *Konets – “Delu” venets (Promezhutochnyy tekstologicheskii finish v bakhtinskom voprose, ili Eshche raz ob avtorstve M. M. Bakhtina v “spornyykh tekstakh”)*, „Dialog. Karnaval. Khronotop“, 2000, № 1, s.72-109 (Пешков Игорь: *Конец — «Делу» венец (Промежуточный текстологический финиш в бахтинском вопросе, или Еще раз об авторстве М. М. Бахтина в „спорных текстах“)*, „Диалог. Карнавал. Хронотоп», 2000, № 1, с.72-109).

- Savoy Jacques: *Is Starnone really the author behind Ferrante?* „Digital Scholarship in the Humanities“ 2018, v. 33, issue 3, 1 September, p.902-918, <https://doi.org/10.1093/llc/fqy016>.
- Skorinkin Daniil, Bonch-Osmolovskaya Anastasiya: *“Osobyue primety” v rechi khudozhestvennykh personazhey: kolichestvennyy analiz dialogov v “Voyne i mire” L. N. Tolstogo*, „Istoriya“, 2016, t. 7, № 7 (51) (Скоринкин Даниил, Бонч-Осмоловская Анастасия: *“Особые приметы” в речи художественных персонажей: количественный анализ диалогов в “Войне и мире” Л. Н. Толстого*, „История“, 2016, т. 7, № 7 (51)).
- Tello José: *What does Delta see inside the Author?: Evaluating Stylo-metric Clusters with Literary Metadata*, in: *III Congreso de La Sociedad Internacional Humanidades Digitales Hispánicas: Sociedades, Políticas, Saberes. Málaga*: HDH, pp.153-161, <http://hdh2017.es/wp-content/uploads/2017/10/Actas-HDH2017.pdf>  
<https://github.com/nevmenandr/ngrams-compare-style/blob/master/fiction/15000/tables/partition2gram.tsv>  
<https://cloud.mail.ru/public/F857/k5M8WooFy>  
<https://cloud.mail.ru/public/Eiae/ZPLQvLLje>  
<https://cloud.mail.ru/public/9pA4/7pqjbpFeX>  
<https://cloud.mail.ru/public/6LAJ/HNEmGhS6v>  
<https://github.com/nevmenandr/ngrams-compare-style/tree/master/take12.2018/nonfiction2/15000/tables>