

МИР ШОЛОХОВА

НАУЧНО-ПРОСВЕТИТЕЛЬСКИЙ
ОБЩЕНАЦИОНАЛЬНЫЙ ЖУРНАЛ

№ 1(11), 2019

МШ

Журнал учреждён Государственным музеем-заповедником М.А. Шолохова
(Министерство культуры РФ)

МИР ШОЛОХОВА
Научно-просветительский общенациональный
журнал

SHOLOKHOV'S WORLD
Scientific and educational national
journal

РЕДАКЦИОННЫЙ СОВЕТ

Г.С. Ермолаев (Принстон, США)
Н.В. Корниенко (Москва)
Богдан Косанович (Нови Сад, Сербия)
В.Н. Крупин (Москва)
В.Я. Курбатов (Псков)
Фредерик Листван (Кельце, Польша)
И.Г. Минералова (Москва)
П.В. Палиевский (Москва)
Л.П. Разогреева (зам. председателя совета,
ст. Вёшенская)
А.М. Шолохов (председатель совета,
ст. Вёшенская)

EDITORIAL COUNCIL

G.S. Ermolaev (Princeton, United States)
N.V. Kornienko (Moscow, Russia)
Bogdan Kosanovic (Novi Sad, Serbia)
V.N. Krupin (Moscow, Russia)
V.Y. Kurbatov (Pskov, Russia)
Frederic Listvan (Kielce, Poland),
I.G. Mineralova (Moscow, Russia)
P.V. Paliyevsky (Moscow, Russia)
L.P. Razogreeva (Deputy President
of the Council, Vyoshenskaya)
A.M. Sholokhov (Chairman
of the Council, Vyoshenskaya)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Главный редактор — **Ю.А. Дворяшин**,
д. фил. н., проф.,
заслуженный деятель науки РФ
Шеф-редактор — **А.А. Дырдин**,
д. фил. н., проф.
Ответственный секретарь — **С.А. Васильев**,
д. фил. н., проф.
Отд. эстетики и культурологии — **Е.А. Костин**,
д. фил. н., проф.
(Вильнюс, Литва)
Отд. литературоведения — **П.С. Глушаков**,
д. фил. н., проф. (Рига, Латвия)
Отд. языкознания — **Л.Б. Савенкова**,
д. фил. н., проф.,
Давыдова О.А., к. фил. н., доц.
Отд. краеведения и музейного дела —
Т.О. Осипова, к. фил. наук, доц.

EDITORIAL BOARD

Editor-in-chief — **Yu. A. Dvoryashin**,
Doctor of Philology, Prof.,
Honored Researcher of Russia (Moscow)
Deputy Chief Editor — **A.A. Dyrdin**,
Doctor of Philology, Prof.
Executive Secretary — **S.A. Vasilev**,
Doctor of Philology, Prof.
Dep. Aesthetics and Cultural Studies —
Y.A. Kostin, Doctor of Philology, Prof.
(Vilnius, Lithuania)
Dep. of Study of Literature — **P.S. Glushakov**,
Doctor of Philology, Prof. (Riga, Latvia)
Dep. of Linguistics — **L.B. Savenkova**,
Doctor of Philology, Prof.,
Davydova O.A., Cand. of Philology, As. Prof.
Dep. of Ethnography and Museum Studies —
T.O. Osipova, Candidate of Philology, As. Prof.

Редактор-корректор — **А.Г. Васильева**
Дизайн обложки — **М.А. Кузнецов**
Макет и вёрстка — **М.Л. Красильникова**

Editor-corrector — **A.G. Vasilyeva**
Design of a cover — **M.A. Kuznetsov**
Layout and typesetting of **M.L. Krasilnikova**

Журнал представлен в Научной электронной библиотеке (НЭБ) e-library.

Мнения авторов могут не совпадать с точкой зрения редакционной коллегии. Редакция журнала не имеет возможности рецензировать и возвращать присланные рукописи, а также вступать в полемику по поводу материалов, не принятых к публикации.

При перепечатке материалов обязательно соблюдение авторских прав и ссылка на публикацию в журнале «Мир Шолохова».

Произведения М.А. Шолохова цитируются по изданию: *Шолохов М.А. Собр. соч.: в 8 т. М.: Художественная литература, 1985–1986, с указанием тома римскими, а страниц — арабскими цифрами.*

Адрес редакции: 346270, Россия, ст. Вёшенская, ул. Шолохова, 60.
E-mail: mirsholokhova@yandex.ru; веб-сайт: <http://www.sholokhov.ru>

Журнал зарегистрирован в Федеральной службе по надзору в сфере связи и массовых коммуникаций. Свидетельство о регистрации СМИ — ПИ № ФС 77-57953 от 28.04.2014.

© Государственный музей-заповедник М.А. Шолохова, 2019.

© Редакция журнала «Мир Шолохова». Название, состав, оформление, редактирование материалов, подготовка текстов к печати, 2019.

Отпечатано в ГУП МО «Коломенская типография».
Подписано в печать 30.04.2019. Формат. 60×90/8. Усл.-печ. л. 22.
Тираж 200 экз. Заказ 632.

Содержание

А.М. Шолохов. Судьба писателя — судьба Родины.	5
Л.П. Разогреева. «Роман под открытым небом» (Государственному музею-заповеднику М.А. Шолохова — 35 лет)	6
ЛИТЕРАТУРОВЕДЕНИЕ	
Н.В. Корниенко. Фольклор и Шолохов (К публикации статьи И.И. Кравченко)	12
И.И. Кравченко. Шолохов и фольклор	29
Н.П. Великанова, Б.В. Орехов. Цифровая текстология: атрибуция текста на примере романа М.А. Шолохова «Тихий Дон»	70
Л.Б. Савенкова. Материнство как ценность в раннем творчестве Шолохова	83
А.В. Дмитрова. Поэтика чёрного цвета в романе Михаила Шолохова «Тихий Дон»	97
Е.А. Ширина. Смех и героика в романе М.А. Шолохова «Они сражались за родину»	102
КЛАССИКА ШОЛОХОВЕДЕНИЯ	
А.А. Горелов. «Тихий Дон» М. Шолохова и русское народное поэтическое творчество (окончание)	107
ФИЛОЛОГИЧЕСКИЕ ШТУДИИ	
Н.Н. Заварзина. Шолоховские мотивы в творчестве Николая Корсунова	123
Н.В. Тибушкина. «Осенняя» образность в романе М.А. Шолохова «Тихий Дон»	128
ИСКУССТВО	
О.И. Бахтиярова. Мастерская Лидии Шабаршиной	134
ПЕДАГОГИЧЕСКОЕ ТВОРЧЕСТВО	
С.А. Васильев. «В одну телегу впрячь не можно / Коня и трепетную лань»?	137
ИСТОРИЯ. КРАЕВЕДЕНИЕ. ЭТНОГРАФИЯ	
Н.И. Фокин. Традиционная казачья семья. Фрагменты из повести-воспоминания «Мы родились и жили на Урале реке...»	148
IN MEMORIA	
Памяти Н.И. Стопченко	169
Памяти Г.С. Ермолаева	171
Об авторах	172
Резюме	175

Н.П. Великанова, Б.В. Орехов

Цифровая текстология: атрибуция текста на примере романа М.А. Шолохова «Тихий Дон»

В статье рассматриваются возможности развития цифровой текстологии, анализируются системные причины применения методов количественной атрибуции авторского текста при решении текстологических задач на примере романа М.А. Шолохова «Тихий Дон». Среди них называются мифологизация статистических подходов и недостаточность проверочной выборки. В качестве решения предлагается надёжный метод атрибуции Delta, разработанный в начале 2000-х годов и применяемый для текстов на множестве языков.

Delta, результаты которой можно представить в виде дендрограммы, хорошо различает произведения современников Шолохова, определяя стиль разных авторов.

В приложении к проблеме авторства «Тихого Дона» Delta хорошо отличает роман от текстов основных претендентов на авторство (Крюкова, Севского, Серафимовича), и результаты её работы бесспорны: все части романа «Тихий Дон» написаны одним автором, причём тем же, кто написал «Донские рассказы», публикуемые под именем Шолохова. Таким образом, именно Шолохов оказывается единственным вероятным автором «Тихого Дона» из всех советских писателей 1920–1930 гг., тексты которых оцифрованы и доступны для компьютерного исследования.

Ключевые слова: Шолохов; «Тихий Дон»; авторство; авторский текст; цифровая текстология; оцифрованный текст; атрибуция; цифровые методы; количественные методы; Delta.

Текстология как наука, генерирующая точные и объективные знания о текстах классического наследия, за свою почти вековую историю претерпела множество трансформаций под влиянием научных, политических, культурных и даже социальных ориентиров того или иного времени. Между тем каждый новый цикл её развития, как ни парадоксально, с одной стороны, добавлял очередную порцию неопределённости и субъективности полученного результата, а с другой — открывал новые возможности, реализация которых зависела от выбора целей, материала и методов исследования, которые разрабатывались и в смежных областях знания, в том числе в истории, философии, искусствознании, лингвистике, социологии, математике, статистике и т.д.

Развитие методов компьютерной лингвистики создало предпосылки для формирования направления, которое можно назвать *цифровой текстологией* [3], предусматривающей получение нового знания при извлечении и обработке информации из оцифрованных рукописных и печатных текстов с применением цифровых методов и инструментов. Таким образом, появилась возможность проверить гипотезы предыдущих циклов развития текстологии, которые не получили статуса достоверного знания, в особенности в категориях «датировки» —

определения историко-хронологической идентичности — и «атрибуции» — определения авторства текста.

Определение авторства текста представляется одной из самых основных и сложных задач текстологического исследования, которая вовсе не сводится к обнаружению плагиата, то есть сознательному присвоению автором чужого текста. И здесь следует разделять три задачи: «прикрепление анонимного или псевдонимного литературного текста к автору» [5], определение соавторства и выявление в тексте «авторского» и «чужого» слова.

Эти задачи в доцифровую эпоху решались эвристически, с применением методов сравнительно-сопоставительного и сравнительно-исторического анализа, изучения документальных свидетельств (писем, дневников, мемуаров), различных способов выявления общих и существенно индивидуальных признаков стиля и языка писателя, их динамики, инструментов систематизации и классификации. В результате выдвигались гипотезы, которые, как правило, вызвали дискуссии, а их убедительность зависела от авторитета и таланта исследователя.

Следует отметить, что выбор методов решения этих задач зависел от целей и объекта исследования, то есть определения авторства целого произведения, его части или отдельных элементов. Причём решение третьей задачи, выявления авторского и неавторского слова, предусматривало, с одной стороны, обнаружение чужого вмешательства, искажающего авторский смысл (цензура, переписчики, редакторы), а с другой — определение намеренного использования автором «чужого» слова (прямые и скрытые цитаты, реминисценции).

Если же целью атрибуции текста является установление факта плагиата, необходим комплексный подход, то есть решение всех трёх задач на основе анализа разночтений всех имеющихся рукописных и печатных источников, описания истории создания и публикации текста.

В случае с определением авторства романа «Тихий Дон» ни одно исследование не продемонстрировало даже возможность такого комплексного подхода. Именно поэтому выводы оказались столь различны: одни считали авторство М.А. Шолохова бесспорным, другие его отрицали полностью, третьи пытались доказать соавторство с разными писателями, четвёртые обнаруживали тематический и сюжетный плагиат, пятые отыскивали лексические и стилистические заимствования.

Между тем в научном издании романа [20], в котором не ставилась задача определения авторства текста, был осуществлён трудоёмкий комплексный анализ текста по всем имеющимся источникам, включая рукописи и печатные издания всех частей романа. В послесловии к роману [6, 7] последовательно и подробно описаны истоки и этапы творческой истории и генезис текста, исторические источники, к которым, как и Л.Н. Толстой при работе над книгой «Война и мир», обращался М.А. Шолохов, социально-исторический, политический, литературный и творческий контексты, вмешательства цензуры и редакторов при публикации, изданиях и переизданиях отдельных частей и романа в целом, ошибки переписчиков и наборщиков, неоднозначное восприятие читателей и критиков и т.д. Немало сил было потрачено и на выявление «чужого» слова, изменявшего авторский смысл.

Именно такое исследование, проведённое на большом массиве текстов и источников, оказалось аргументированным, доказательным и убедительным, правда, недоступным широкому кругу читателей в силу того, что издание выпущено небольшим тиражом.

Первое научное издание текста романа «Тихий Дон», очищенного от неавторских искажений, стало основанием проверки старых гипотез авторства текста новыми цифровыми инструментами, апробированными на множестве зарубежных и русских текстов и показавшими убедительный результат.

История количественных исследований текста романа «Тихий Дон»

Для цифровых методов в отечественной филологии проблема авторства романа «Тихий Дон» стала своего рода традиционной областью применения. Такие же дежурные «полигоны» апробации методов есть и в других научных традициях. Например, для англоязычного мира это тексты Шекспира и Дж. Остин (в последнем случае речь идёт уже не об атрибуции, а о максимально широком круге задач).

Высокая общественная значимость этой темы продолжает активно привлекать внимание исследователей, знакомых с математическим аппаратом, а в последнее время — и с компьютерными инструментами. Среди последних работ на эту тему следует назвать статьи В.А. Бубнова и А.В. Сурвило [1; 2], М.Ю. Михеева и Л.И. Эрлиха [14], а также только что опубликованное исследование М. Иосифян и И. Власова [29].

Однако ситуация всё новых обращений вооружённых точными методами учёных к этой проблематике сложилась далеко не вчера, и у этого есть ряд системных причин. Сама структура аргументации в дискуссии вокруг донской эпопеи такова, что вес любого факта или утверждения зависит от предзаданной позиции исследователя. Если в силу политической или эстетической ангажированности исследователь склонен придерживаться той или иной точки зрения на проблему авторства, то любое общее соображение, конкретное наблюдение или биографическое свидетельство он будет трактовать как подтверждающее его взгляды. Этот психологический эффект хорошо описан в англоязычной литературе и ёмко называется *confirmation bias* [36] (удачного перевода на русский язык пока не придумано, обычно говорят «склонность к подтверждению своей точки зрения» или «предвзятость подтверждения»). О нём же в случае с близкой к атрибуции проблемой датирования «Слова о полку Игореве» писал и А.А. Зализняк: «Очень часто из предъявляемого факта решительно ничего не вытекает с обязательностью, и даже, если взглянуть на тот же самый факт чуть иначе, он начинает выглядеть как свидетельство в пользу противоположной версии. Поэтому, как это ни поразительно, сторонники противоположных точек зрения нередко ссылаются на одни и те же факты» [8, 20]. Действительно, если для занимающих одну позицию в споре наличие рукописи «Тихого Дона» служит неоспоримым аргументом авторства М.А. Шолохова, то другие видят в этом лишь часть циничного плана плагиата. Так же условны и неубедительны другие факты и свидетельства, которые вовлекаются в орбиту дискуссии.

В этой ситуации кажется, что цифровые методы должны защитить научный строй мысли от ангажированности и стать более весомым аргументом, чем ненадёжные соображения, да ещё и трактуемые через призму эмоций.

Два самых известных высказывания в этом поле — монография скандинавских исследователей [18] и исследование А.Т. Фоменко [17]. В обоих случаях авторы обратились к подсчётам, чтобы выйти из замкнутого круга *confirmation bias* и попробовать добавить в дискуссию объективные ориентиры.

Научный анекдот состоит в том, что исследователи получили противоположные результаты. Несмотря на декларированный подход, основанный на бесстрастной математике, попытка решения вопроса об авторстве романа не только не привела к искомому результату, но и создала неприятную ситуацию инфляции доверия к количественным аргументам в научной дискуссии. В самом деле, если в теории подсчёты должны оградить нас от влияния политики и эмоций, но на практике допускают противоположные трактовки, может быть, использующие их авторы точно так же оказываются в плену своих взглядов и просто ищут в цифрах подтверждения априорной убеждённости? Тогда количественные методы атрибуции выглядят ничуть не лучше, а может быть, даже менее выигрышно по сравнению с традиционными аргументами. Ведь в отличие от биографических, текстологических и прочих подобных свидетельств, хорошо понятных гуманитариям, количественные методы оказываются своего рода «чёрным ящиком»: им на вход подаётся исследовательский вопрос и в результате непонятных действий внутри на выходе появляется ответ, степень обоснованности которого неизвестна. Не имеющему специальной подготовки филологу трудно оценить, почему в результатах разных исследователей получаются такие расхождения, какие подсчёты сделаны правильно, а какие ошибочно.

Однако нам представляется, что дело тут вовсе не в математике. То, что в конечном счёте приводит к девальвации метода количественной атрибуции, кроется в самом построении аргументации применяющих её исследователей.

Если упрощённо формулировать общий подход к определению авторства романа «Тихий Дон», то средняя работа выглядит так: «Мы подсчитали параметр X в некоторых текстах, и это должно значить следующее...»

Так поступил, например, Г. Хьетсо: «Средняя длина предложений у Крюкова составляет 13,9 слова, у Шолохова — 12,9 слова, тогда как в “Тихом Доне” — 12,4 слова. Таким образом, можно сказать, что произведения Шолохова обнаруживают в этом смысле больше сходства с “Тихим Доном”, чем произведения Крюкова» [18, 61]. Однако никаких практических гарантий того, что средняя длина предложений вообще является релевантным для атрибуции параметром, нам не даётся. Ещё едва ли не в самой первой работе с использованием точных методов определения авторства Т.К. Менденхолл пытался доказать, что авторский сигнал в тексте выражен в средней длине слов [34]. Но позднее обнаружилось, что этот параметр никак не зависит от стиля автора, и подсчитывать его не имеет смысла.

Иными словами, исследователи, использующие точные методы, предлагают верить их выводам только на том основании, что те базируются на инструментах математической статистики. Такой подход в достаточной мере мифологичен. Логическая аргументация заменяется риторической фигурой, эксплуатирующей венауочное представление о безусловной правильности всего, что связано с математикой. В последнее время оно трансформировалось в миф о «компьютерном анализе» [15], который, кстати, задействовал и Г. Хьетсо, специально подчёркивая в предисловии, что работе предшествовала «компьютерная обработка» [18, 13]. Почему-то упоминание компьютерного анализа воспринимается читателем как более убедительное. Однако, по большому счёту, на качество выводов не должно влиять то, как производились подсчёты, — с использованием мощного компьютера или карандаша и бумаги.

А.Т. Фоменко к решению задачи определения авторства романа «Тихий Дон» подошёл с точки зрения общей методологии «обнаружения “авторского инвари-

анта” для русских литературных текстов» [17, 688], стремясь обосновать корректность своего метода на примере анализа текстов разных писателей, и уже затем приступил к решению задачи атрибуции «Тихого Дона». В исследовании отражено понимание того, что количественная оценка индивидуальных особенностей автора — задача весьма сложная. Поэтому авторский инвариант должен обладать тремя свойствами: «бессознательным параметром», постоянным значением для произведений автора, различающей способностью от других писателей. Кроме того, выборка должна включать не менее 16 000 слов. Важно отметить, что исследование больших массивов текстов 28 русских и советских писателей проводилось в 1970-е годы вручную небольшой группой учёных, и в результате эксперимента был определен авторский инвариант — «количество всех служебных слов (в процентах)» [17, 704].

Найденный авторский инвариант для русских литературных текстов А.Т. Фоменко использовал для определения авторства романа «Тихий Дон» на основе сравнительно-статистического анализа произведений М.А. Шолохова и Ф.Д. Крюкова.

Такая последовательность в большей степени соответствует научному подходу, хотя в итоге исследователи остались заложниками своего времени и возможностей, так как не смогли проверить выкладки на значимо большей текстовой выборке произведений советских писателей, написанных в период писания и печатания «Тихого Дона», в том числе и расшифрованных рукописей первой и второй книг романа «Тихий Дон» [22], которые были обнаружены ещё в прошлом веке (об этом А.Т. Фоменко не знал) и опубликованы в факсимильном издании в 2005 г. [21]. В результате эксперимента получилось, что «Донские рассказы», роман «Поднятая целина», поздние произведения, а также значительная часть третьей книги и четвёртая книга романа «Тихий Дон» написаны М.А. Шолоховым, в то время как книги первая и вторая, а также начало книги третьей донской эпопеи автору «Донских рассказов» не принадлежат.

В продолжение эксперимента А.Т. Фоменко математик А.В. Зенков предложил ещё один статистический метод, который, как ему представлялось, позволяет ответить на вопрос об авторстве текста, в частности романа «Тихий Дон». Подсчитав первые значащие цифры числительных, исследователь пришёл к выводу о том, что «в высшей степени сомнительно, что “Донские рассказы”, с одной стороны, и “Тихий Дон”, “Поднятая целина”, “Они сражались за родину” принадлежат одному автору» [9, 558], тем самым полностью отказав М.А. Шолохову в авторстве «Тихого Дона».

Так формируется негативный имиджевый фон количественной атрибуции, основной предпосылкой которого становится непродуманность аргументации и недостаточная для проверки гипотезы текстовая выборка.

Следует сказать, что богатая отечественная традиция, в которой предлагаются различные точные методы атрибуции [напр.: 11; 10; 19], так же упирается в слабую доказательную базу, не предлагая подтверждения работоспособности методов на достаточно обширном проверочном материале.

Как работает метод Delta

Итак, главная проблема предыдущих попыток количественного ответа на вопрос об авторстве «Тихого Дона» состояла в том, что исследователи изобретали свой метод атрибуции до или в процессе доказательства. Не проверив его и

спеша подтвердить свою точку зрения по поводу авторства Шолохова, они попадали в ловушку логического круга. Учёным приходилось одновременно решать сразу две задачи: и разработки метода, и атрибуции «Тихого Дона», что можно сравнить с войной на два фронта, а это, как мы знаем из истории, является крайне уязвимой и малоперспективной ситуацией.

Если же нас интересует именно проблема авторства донской эпопеи, то правильный шаг состоит не в том, чтобы изобретать свой метод атрибуции, вполне возможно, не работающий, а в том, чтобы воспользоваться существующими авторитетными инструментами, которые дают гарантированный результат.

На момент, когда создавались труды Г. Хьетсо с соавторами и А.Т. Фоменко, таких методов действительно не было. Но текущая ситуация кардинально отличается тем, что такой метод атрибуции существует, и работает он настолько хорошо, что мы можем констатировать: задача количественной атрибуции в настоящее время успешно решена.

Речь идёт о методе, который называется Delta [25]. Появившись в начале 2000-х гг., он приобрёл широкую популярность и признание, многократно подтвердив свою эффективность для разных языков, например арабского [23], польского [26], немецкого [32], древнеанглийского [27], в том числе и для русского [16]. Только в течение 2017–2018 гг. вышло не меньше десятка статей [напр.: 24; 33; 35; 31], в которых Delta успешно используется для решения разных стилеметрических задач, в первую очередь, конечно, связанных с проблемами установления авторства.

Delta работает следующим образом. Для каждого из анализируемых текстов берётся некоторое заданное количество самых частотных слов, а затем сравниваются их частотности (представленные как особая мера *z-score*, учитывающая, например, число всех слов в тексте), суммируя разницу (ту самую *дельту*) для каждой пары текстов. Эмпирическая закономерность заключается в том, что разница для текстов, написанных одним автором, оказывается меньше, чем для любой пары текстов, написанных разными людьми.

Расстояние между всеми текстами в анализируемой выборке можно было бы представить в виде таблицы, но, кажется, это не очень наглядно и требует от читателя дополнительных усилий по выискиванию в ней нужных значений. Поэтому в стилеметрии с некоторых пор активно применяется такой тип визуализации, как дендрограмма (рис. 1).

Читать дендрограмму нужно следующим образом. Мы взяли 32 текста писателей 1920–1930-х годов, то есть периода публикации частей романа «Тихий Дон». При этом первую и вторую части романов Островского «Как закалялась сталь» и Фадеева «Последний из Удэге» мы представили в виде отдельных единиц корпуса (далее мы проследим, не путает ли Delta их с текстами других писателей), а роман Серафимовича «Железный поток» разделили примерно пополам: в первой половине оказались главы с 1-й по 22-ю включительно, во второй – все остальные (сможет ли Delta определить, что это произведение одного автора?). Мы померили расстояние между единицами корпуса с помощью Delta, используя 200 самых частотных слов, общих для всех этих текстов. Далее стилистическое сходство текстов было представлено в виде дерева. При отсчёте от правой части изображения на каждом шаге самые стилистически сходные тексты объединяются в одну ветку. Похожие произведения оказываются сгруппированы вместе, а самые непохожие (но сходные между собой) – относятся в другую часть рисунка. Так, видно, что тексты М. Горького силь-

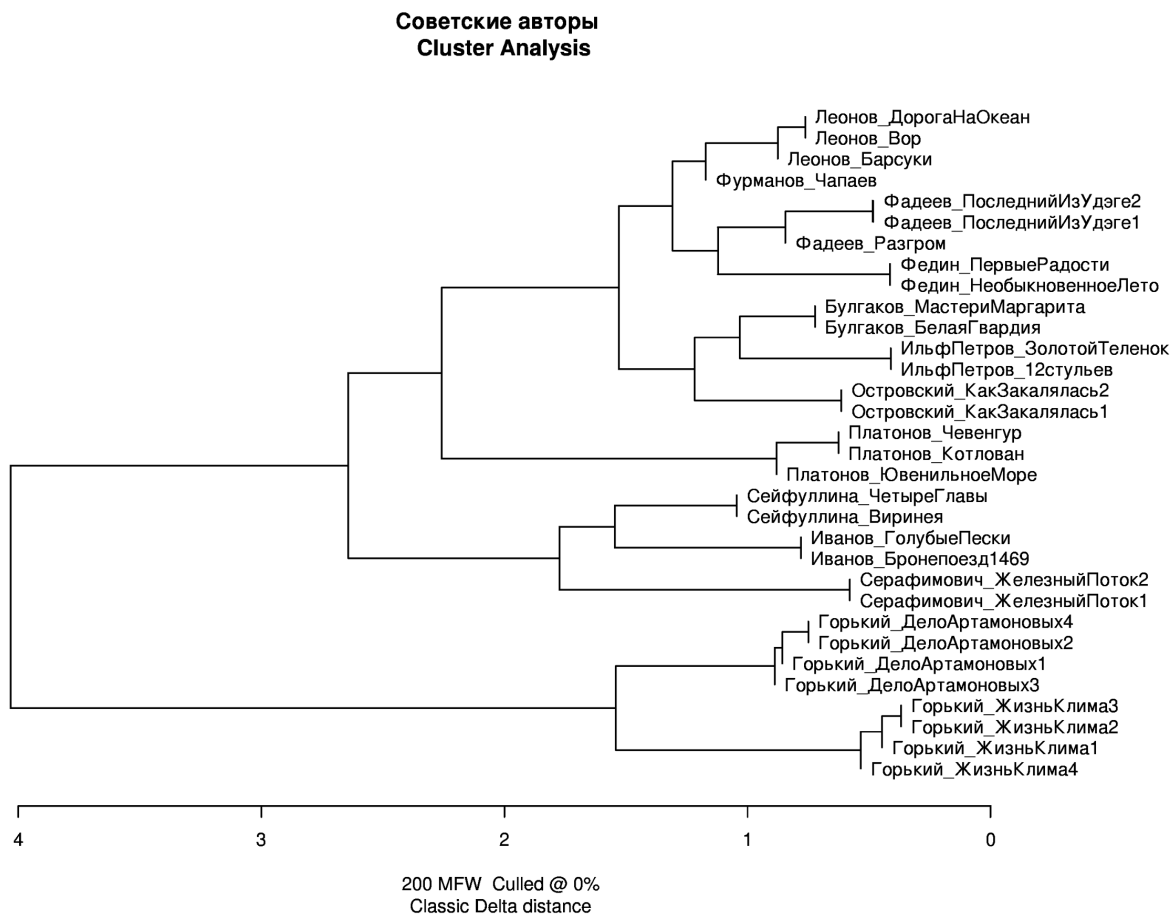


Рис. 1. Дендрограмма, визуализирующая расстояние Delta для текстов советских писателей

но отличаются от всех остальных. Произведения Серафимовича, Вс. Иванова, Сейфулиной составляют отдельную группу, которая объединяется с остальными только на предпоследнем шаге объединения. На третьем с конца шаге с остальными сходятся произведения Платонова. На четвёртом с конца шаге одну группу составляют романы Леонова, Фурманова, Фадеева и Федина, а другую — Булгакова, Ильфа и Петрова, Островского. Наконец, на первом шаге (в правой части изображения) самые похожие тексты оказываются «листами» на соседних ветвях дерева.

Как мы видим, Delta безошибочно определила, какие произведения написаны одним автором. Тексты Вс. Иванова оказываются на соседних ветвях, и это означает, что именно между ними обнаружилось минимальное расстояние. Аналогичную ситуацию мы наблюдаем для каждого автора, который представлен в корпусе более чем одним текстом. Только Д. Фурманов представлен единственным романом «Чапаев», но мы видим, что это произведение не путается с текстами других авторов, то есть не попадает, например, на ветку между романами Леонова или Федина. Романы М. Горького составляют в нашем дереве отдельную большую ветвь. Примечательно, что Delta не только сумела отделить Горького от остальных авторов, но и выявить стилистическую разницу между романами «Дело Артамоновых» и «Жизнь Клим Самгина»: разные части одного романа оказываются стилистически похожи между собой больше, чем части другого романа того же автора.

Разумеется, у метода есть свои ограничения. Так, надёжные результаты с его помощью могут быть получены только на жанрово гомогенном корпусе, а каждый составляющий этот корпус текст должен быть по объёму не меньше 10 000 слов. Тем не менее, результаты Delta несомненны и многократно подтверждены. По большому счёту, исследователю, имеющему этот метод в своём арсенале, совершенно не нужно искать другие способы количественной атрибуции для стандартных задач. Проблемы могут возникнуть только там, где требуется решить вопрос авторства в ситуации доступности ограниченного круга коротких или разножанровых текстов. Но все эти сложности не касаются случая Шолохова и «Тихого Дона». Кроме того, важно, что поскольку в подсчётах учитываются только частотные слова, тематика текста не влияет на точность атрибуции: тему задают значащие части речи (существительные, прилагательные и глаголы), частотность каждого из которых несопоставимо низкая по сравнению с частотностью практически любого предлога или союза.

Как анализировался текст «Тихого Дона»

Теперь обратимся к тексту романа «Тихий Дон». Возьмём для нашего корпуса 13 произведений семи советских писателей, добавим по отдельности четыре книги «Тихого Дона» нового, научного издания [20], «Донские рассказы», поздние произведения Шолохова (включая «Судьбу человека», так как в этом рассказе чуть более 10 000 слов, порогового значения, ниже которого Delta даёт ненадёжные результаты) и тексты трёх кандидатов на авторство романа — Ф.Д. Крюкова, Виктора Севского и А. Серафимовича.

Главным претендентом, конечно, является Ф.Д. Крюков (1870—1920), кандидатура которого возникла ещё в 1937—1938 гг., когда роман не был завершён и читатели ждали выхода в свет последней книги. Эта кандидатура затем была предложена в 1974 г. И.Н. Медведевой-Томашевской, женой известного филолога Б. Томашевского, которая пыталась доказать двойное авторство всего романа, основываясь на вероятностном анализе сюжетных линий и противоречиях в описании исторических реалий. Её книга «Стремя “Тихого Дона” (Загадки романа)» [13] при содействии и с предисловием А.И. Солженицына была напечатана парижским издательством YMCA-PRESS и спровоцировала волну изысканий, подтверждающих или опровергающих авторство М.А. Шолохова, в том числе и идеологически мотивированных, как, например, книга историка Р. Медведева [30]. В 1980-е годы открылась возможность для обсуждения проблемы авторства «Тихого Дона» и появилась новая волна публикаций, в которых авторство приписывалось Ф.Д. Крюкову, преимущественно на основе неподтверждённых биографических фактов.

В корпус исследования мы включим пять наиболее объёмных текстов из наследия Ф.Д. Крюкова.

Израильский филолог Зеев Бар-Селла (ранее В.П. Назаров) настаивал на том, что авторство «Тихого Дона» принадлежит журналисту В.А. Краснушкину (псевдоним «Виктор Севский»), который, как утверждал автор гипотезы, был казаком, противником большевиков, свидетелем гражданской войны на Дону, вел дневник, много печатался в газетах, а поэтому вполне мог написать «Тихий Дон». Нам доступно одно его произведение достаточного объёма.

М. Марусенко отдал авторство донской эпопеи А. Серафимовичу [12], текст которого, к счастью, уже представлен в нашем корпусе. Эта версия, кажется,

наиболее экстравагантная из всех, ведь именно А. Серафимович в конце 1920-х годов возглавлял комиссию, которая должна была расследовать подозрения в плагиате М.А. Шолохова. Получается, что А. Серафимович должен был разбираться с украденным у него же самого романом.

Существуют и другие предположения относительно возможных авторов романа, художественные тексты которых нам недоступны и не могут быть включены в настоящее рассмотрение.

Так, А.А. Венков, профессор истории, указывал на фигуру военного чиновника И.Д. Филиппова [4], который выступил соавтором диалогов и любовных линий сюжета первых частей романа. Произведения И.Д. Филиппова, печатавшиеся в журналах «Голос казачества» и «Вестник казачества», можно будет учесть после их оцифровки. А.А. Венков также утверждал, что сцены военной жизни представляют собой переработанные публикации донских офицеров в газетах и журналах того времени.

С.С. Голоушев (псевдоним «Сергей Глаголь»), автор путевого очерка «С Тихого Дона» (1917), имя которого в связи с возможным плагиатом Шолохова упоминалось ещё в 1930-х годах, оставил в основном критические произведения, которые невозможно адекватно соотнести с художественными текстами.

Кроме того, следует учесть и многочисленные версии того, что Шолохов является автором не всей книги, а только последних частей, а также не является автором и других произведений, печатавшихся под его именем. Так, «Поднятая целина» приписывается К. Каргину, которому принадлежит рассказ «Забурунный край», к сожалению, нам недоступный.

Результаты применения к собранному корпусу метода Delta представлены на рис. 2.

Прежде всего мы видим, что Delta снова не допустила ошибок в бесспорных случаях. Нет сомнений в том, что «Мастер и Маргарита» и «Белая гвардия» написаны одним автором, и Delta показала минимальное расстояние между ними: оба текста помещены на визуализации на соседние ветви. То же самое произошло с уже знакомыми нам по рис. 1 произведениями Н. Островского, Л. Леонова, А. Серафимовича и Вс. Иванова, а также введенными в обозреваемый материал на этом этапе рассказами Ф. Крюкова. Это говорит о том, что Delta продолжает работать исправно, и у нас есть основания доверять полученным результатам.

Некоторая странность наблюдается только в том, что «Донские рассказы» и поздние тексты Шолохова оказались в разных частях дерева. В то же время это вполне естественно: роман писался 16 лет, стиль автора за это время должен был меняться. Во всяком случае, «Поднятая целина», «Они сражались за родину» и «Судьба человека» не оказались разбросаны по разным ветвям, а это означает, что нет причин подозревать для них разных авторов: если К. Каргин действительно написал «Поднятую целину», то его же перу должны принадлежать и другие поздние тексты, подписанные именем Шолохова.

Все книги «Тихого Дона» так же, по всей видимости, написаны одним человеком, что противоречит многочисленным попыткам антишолоховедов выделить в них инородные вставки, разграничить «оригинальный» текст и исправления «соавторов».

Наиболее существенно, что, по данным Delta, автор «Тихого Дона» должен одновременно быть и автором «Донских рассказов», причём больше всего на них

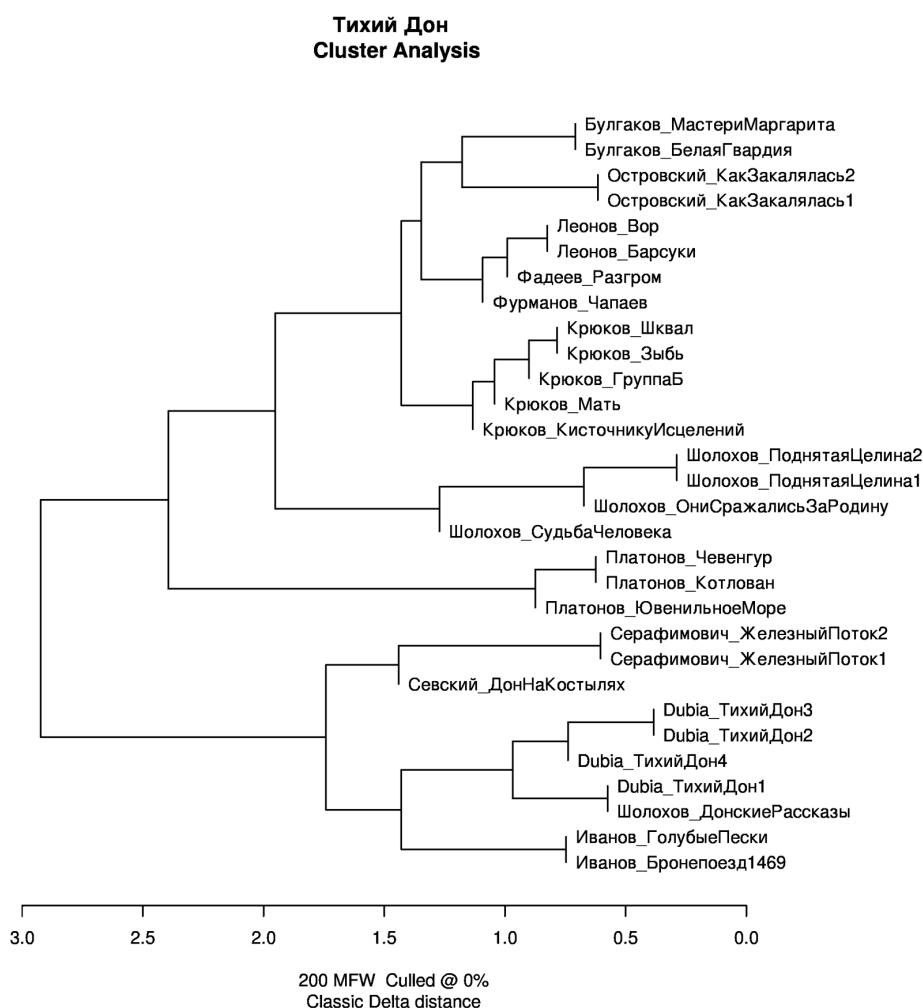


Рис. 2. Дендрограмма расстояний для текстов претендентов на авторство «Тихого Дона»

похож именно первый том романа, который чаще всего приписывают не Шолохову, а какому-либо кандидату из пула претендентов на авторство, оставляя за Шолоховым право претендовать на финальную часть эпопеи. Такое представление ситуации маловероятно: первый том «Тихого Дона» и «Донские рассказы» писал один и тот же человек и, как известно, в одно и то же время, во второй половине 1920-х годов.

Delta продемонстрировала, что уверенно различает стилистику Ф. Крюкова, расположив все его произведения листьями на соседних ветвях. И так же уверенно можно сказать, что Ф. Крюков не писал роман «Тихий Дон», поскольку дендрограмма уверенно отделяет его тексты от романа Шолохова и помещает в одном корпусе с текстами Булгакова, Островского, Леонова, Фадеева, Фурманова и Платонова, причём в этом же корпусе оказались все остальные произведения Шолохова — «Поднятая целина», «Судьба человека», «Они сражались за родину», авторство которых Крюкову не приписывается.

Авторство В. Севского для «Тихого Дона» ещё менее вероятно, чем авторство Вс. Иванова, которого, кажется, пока никто не называл в числе возможных создателей романа. И совсем нет никаких причин подозревать, что «Тихий Дон» написан А. Серафимовичем.

У этой визуализации есть числовые соответствия. Так, Delta-расстояние от «Мастера и Маргариты» до «Белой гвардии» 0,73, то есть очень маленькое, а от «Донских рассказов» до первого тома «Тихого Дона» 0,59, то есть значительно меньше. Ради справедливости надо сказать, что в числах Delta показывает и значительную близость «Они сражались за родину» и последнего тома «Тихого Дона»: 0,72 — эти книги стилистически даже ближе между собой, чем два романа Булгакова.

Итак, главный вывод однозначен — если признавать «Донские рассказы» за Шолоховым, то именно он написал и все части «Тихого Дона». Другие претенденты, включённые в наше рассмотрение, не имеют шансов быть названными авторами романа. Побочный вывод состоит в том, что поздние произведения Шолохова плохо подходят для решения проблемы атрибуции «Тихого Дона»: стиль писателя менялся, тексты подвергались значительной правке редакторов, и в результате меняется фиксируемый Delta авторский сигнал.

Исследование, проведённое с помощью цифрового метода Delta, показало новые возможности для получения убедительных результатов атрибуции авторского текста. Кажется, что эта важная и сложная задача решена, но есть и другие задачи, которые ждут решения в области цифровой текстологии.

Библиографический список

1. Бубнов В.А., Сурвило А.В. Сравнительный компьютерный анализ текста романа «Тихий Дон» с текстами четырёх рассказов Фёдора Крюкова // Вестник Российского университета дружбы народов. Серия: Информатизация образования. 2016. № 1. С. 60–69.
2. Бубнов В.А., Сурвило А.В. Сравнительный компьютерный анализ текста романа «Тихий Дон» с текстами Фёдора Крюкова, Александра Серафимовича и Ивана Родионова // Вестник Российского университета дружбы народов. Серия: Информатизация образования. 2016. № 2. С. 73–82.
3. Великанова Н.П. «Тихий Дон» М.А. Шолохова: текстологические решения научного издания // Известия Российской академии наук. Серия литературы и языка. 2018. Т. 77. № 6. С. 69–76.
4. Венков А. «Тихий Дон»: источниковая база и проблема авторства. Ростов н/Д.: Терра, 2000. 584 с.
5. Виноградов В.В. Проблема авторства и теория стилей. М.: Художественная литература, 1961.
6. Воронцова Г.Н. Текстологическое послесловие // Шолохов М.А. Тихий Дон. Научное издание: в 2 т. Т. 2. М.: ИМЛИ РАН, 2017. С. 781–855.
7. Дворяшин Ю.А. Текстологическое послесловие // Шолохов М.А. Тихий Дон. Научное издание: в 2 т. Т. 1. М.: ИМЛИ РАН, 2017. С. 703–812.
8. Зализняк А.А. «Слово о полку Игореве»: взгляд лингвиста. М.: Рукописные памятники Древней Руси, 2008.
9. Зенков А.В. Новый статистический метод атрибуции текстов // XI Международная конференция «Российские регионы в фокусе перемен»: Сборник докладов. Екатеринбург: изд-во УМЦ УПИ, 2016. Ч. 1. С. 551–562.
10. Кукушкина О.В., Поликарпов А.А., Хмельёв Д.В. Определение авторства текста с использованием буквенной и грамматической информации // Проблемы передачи информации. М.: Наука, 2001. Т. 37. № 2. С. 96–108.
11. Марусенко М.А. Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов. Л.: ЛГУ, 1990. 164 с.

12. *Мафусенко М.А., Бессонов Б.Л., Богданова Л.М., Аникин М.А., Мясоедова Н.Е.* В поисках потерянного автора: Этюды атрибуции. СПб.: Филологический факультет СПбГУ, 2001. 211 с.

13. *Медведева И.Н. (Д*)*. Стремя «Тихого Дона» (загадки романа). М.: Горизонт, 1993.

14. *Михеев М.Ю., Эрлих Л.И.* Идиостилевой профиль и определение авторства текста по частотам служебных слов // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2018. № 2. С. 25–34.

15. *Орехов Б.В.* Лексическая единица «компьютерный анализ» в языке СМИ // Проблемы современной лингвистики и методики преподавания языковых курсов. Кемерово: Кузбассвузиздат, 2008. С. 143–145.

16. *Скоринкин Д.А., Бонч-Осмоловская А.А.* «Особые приметы» в речи художественных персонажей: количественный анализ диалогов в «Войне и мире» Л.Н. Толстого // Электронный научно-образовательный журнал «История». 2016. Т. 7. Вып. 7 (51).

17. *Фоменко А.Т.* Античность — это средневековье. СПб.: Издательский дом «Нева», 2005.

18. *Хьетсо Г., Густавссон С., Бекман Б., Гил С.* Кто написал «Тихий Дон»? (Проблема авторства «Тихого Дона») / Пер. А.В. Ващенко, Н.С. Ноздриной. М.: Книга, 1989. 186 с.

19. *Хмелёв Д.В.* Распознавание автора текста с использованием цепей А.А. Маркова // Вестник МГУ. Сер. 9: Филология. 2000. № 2. С. 115–126.

20. *Шолохов М.А.* Тихий Дон. Научное издание: в 2 т. / Под ред. Н.В. Корниенко, Г.Н. Воронцовой, Ю.А. Дворяшина, А.Б. Куделина, Ф.Ф. Кузнецова, В.В. Полонского, М.А. Шолохова, А.М. Ушакова, Е.А. Тюриной. М.: ИМЛИ РАН, 2017.

21. *Михаил Шолохов* / Тихий Дон. Рукопись. Книга первая. Книга вторая. Факсимильное издание. Москва—Киев—Париж: ИМЛИ РАН А.М. Горького, «Библиотека журнала «Дніпро»», Московский писатель, 2005.

22. *Михаил Шолохов.* «Тихий Дон». Динамическая транскрипция рукописи. М.: ИМЛИ РАН, 2011.

23. *AbdulRazzaq A.A., Mustafa T.K.* Burrows-Delta Method Fitness for Arabic Text Authorship Stylometric Detection // International Journal of Computer Science and Mobile Computing, Vol. 3. Issue 6, June–2014. P. 69–78.

24. *Alexander A.G., Gladwin Matthew J., Lavin Daniel M.* Look Stylometry and Collaborative Authorship: Eddy, Lovecraft, and ‘The Loved Dead’ // Digital Scholarship in the Humanities. Vol. 32. Issue 1, 1 April 2017. P. 123–140; <https://doi.org/10.1093/llc/fqv026>

25. *Burrows J.* ‘Delta’: a Measure of Stylistic Difference and a Guide to Likely Authorship // Literary and Linguistic Computing. Vol. 17. Issue 3, 1 September 2002. P. 267–287.

26. *Eder M., Rybicki J.* PCA, Delta, JGAAP and Polish Poetry of the 16th and the 17th Centuries: Who Wrote the Dirty Stuff? // Digital Humanities 2009: Conference Abstracts: MD College Park, 2009. P. 242–244.

27. *García A.M., Martín J.C.* Function Words in Authorship Attribution Studies // Literary and Linguistic Computing. 2007. Vol. 22. № 1. P. 49–66;

28. *Hartmut Ilseemann.* Stylometry approaching Parnassus // Digital Scholarship in the Humanities. Vol. 33. Issue 3, 1 September 2018. P. 48–556; <https://doi.org/10.1093/llc/fqx058>

29. *Iosifyan M., Vlasov I.* And Quiet Flows the Don: the Sholokhov–Kryukov authorship debate // *Digital Scholarship in the Humanities*. 2019, fqz017.
30. *Medvedev R.* Problems in the Literary Biography of Mikhail Solokhov. Cambridge: Cambridge University Press, 1977.
31. *Savoy J.* Is Starnone Really the Author behind Ferrante? // *Digital Scholarship in the Humanities*. Vol. 33. Issue 4, 1 December 2018. P. 902–918; <https://doi.org/10.1093/llc/fqy016>
32. *Jannidis F., Lauer G.* Burrows's Delta and Its Use in German Literary History // *Distant Readings. Topologies of German Culture in the Long Nineteenth Century Studies in German Literature Linguistics and Culture* / Eds. M. Erlin, L. Tatlock. Rochester: Camden House, 2014. P. 29–54.
33. *José Calvo Tello.* What does Delta see inside the Author? // *Evaluating Stylometric Clusters with Literary Metadata*. 2018. P. 153–161.
34. *Mendenhall T.C.* The Characteristic Curves of Composition // *Science*. 1887. № 9. P. 237–249.
35. *Michael P. Oakes.* Computer Stylometry of C.S. Lewis's *The Dark Tower* and Related Texts // *Digital Scholarship in the Humanities*. Vol. 33. Issue 3, 1 September 2018. P. 637–650; <https://doi.org/10.1093/llc/fqx043>
36. *Nickerson Raymond S.* Confirmation bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*. 1998. June, 2 (2). P. 175–220.